이학박사 학위논문

Bootstrap in Nonparametric Dynamic Discrete Choice Models for Time Series Data

시계열 자료를 위한 비모수 동적이산선택 모형의 붓스트랩

2019년 8월

서울대학교 대학원

통계학과

주 성 재

Bootstrap in Nonparametric Dynamic Discrete Choice Models for Time Series Data

by

Seongjae Joo

A Dissertation

submitted in fulfillment of the requirement

for the degree of

Ph.D of Science

in

Statistics

The Department of Statistics

College of Natural Sciences

Seoul National University

August, 2019

Abstract

Seongjae Joo

Department of Statistics

The Graduate School

Seoul National University

In this thesis, we investigate bootstrap in nonparametric dynamic discrete choice models. This model allows to involve the lags of the discrete response variable as well as exogenous discrete covariates among regressors. We first propose a methodology for estimation. This is a generaliztaion of local likelihood approach to dynamic discrete choice models. We prove uniform consistency, expansion and asymptotic normality of the estimator. Based on this estimator, we propose two model-based bootstrap procedures. We construct a one-step bootstrap estimator from bootstrap resamples. We show that both bootstrap procedures consistently approximate the laws of the estimator. In the simulation study, we illustrate the performance of bootstrap confidence intervals. We apply dynamic binary probit model to analyze the recession data. We provide estimates of regression function and its pointwise bootstrap confidence intervals.

Keywords: Discrete response models, Nonparametric models, Bootstrap, Time series data

Student Number: 2012-20235

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we investigate bootstrap in nonparametric dynamic discrete choice (response) models for time series data. In our model, the discrete response is related to exogenous regressors which may include both continuous and discrete variables and its own lagged values. Discrete response models have been popular and widely investigated in a variety of areas. Popular applications of time series discrete choice models are the forecasting of recessions, movements, market indices in econometrics and weather forecasting in climatology.

The conditional distribution of the response variable given covariates is modeled by multiparameter local likelihood model. Aerts and Claeskens (1997) investigated multiparameter local likelihood model

$$Y \sim f(\cdot, (\theta_1(\mathbf{x}), \dots, \theta_\nu(\mathbf{x}))) \tag{1.1}$$

where Y is a response variable and f is a known density function involving unknown ν functions $\theta_1, \ldots, \theta_\nu$ depending on a regressor $\mathbf{x}$. They proposed local polynomial estimation of $\theta_1, \ldots, \theta_\nu$ from i.i.d. (independent and identically distributed) sample $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ following (1.1). They applied beta-binomial models to analyze clustered binary data. Park et al. (2015) also dealt with multiparameter local likelihood model involving categorical regressors among covariates. We extend these works to dynamic discrete response models. Park et al. (2017) also considered local linear estimation of dynamic discrete response models. They extended local quasi-likelihood approach of Fan et al. (1995) to the time series models with dynamic features. We follow the methodology for estimation in Park et al. (2017). We employ the likelihood approach because we are studying the model-based bootstrap in which one needs to know the form of the conditional distribution of the response variable given covariates to generate bootstrap resamples. We derive our estimator's properties: uniform consistency, expansion and asymptotic distribution which are required for developing our bootstrap methodology.

The bootstrap is a simple but very powerful methodology to get knowledge from limited data. The bootstrap enables us to assess the uncertainty of our findings. Efron (1979)'s pioneering work increased studies of the bootstrap among statisticians. The bootstrap in linear regression models was investigated by Freedman (1981). Many authors investigated the bootstrap of kernel smoother from i.i.d. observations, e.g., see Hardle and Bowman (1988), Hall

(1992), Neumann (1995), Neumann (1997), Härdle et al. (1998), Neumann and Polzehl (1998), Xia (1998), Claeskens and Keilegom (2003), Härdle et al. (2004) and McMurry and Politis (2008).

The bootstrap for dependent data has also been widely investigated. Hall (1985), Carlstein (1986), Kunsch (1989) and Liu and Singh (1992) introduced block resampling ideas to various statistical inference problems for dependent data. For example, moving block bootstrap of Liu and Singh (1992) divides data $X_1, \ldots, X_n$ into blocks

$$\mathcal{B}_1 = (X_1, \ldots, X_\ell), \mathcal{B}_2 = (X_2, \ldots, X_{\ell+1}), \ldots, \mathcal{B}_{n-\ell+1} = (X_{n-\ell+1}, \ldots, X_n)$$

and randomly draw blocks from $\mathcal{B}_1, \ldots, \mathcal{B}_{n-\ell+1}$. Gonçalves and White (2004) investigated moving block bootstrap in (parametric) dynamic nonlinear models. For nonparametric models, nonlinear autoregression models have been investigated by Neumann and Kreiss (1998), Franke et al. (2002a) and Franke et al. (2002b). Nonlinear autoregression model is formulated by

$$X_t = m(X_{t-1}, \ldots, X_{t-p}) + \sigma(X_{t-1}, \ldots, X_{t-p})\varepsilon_t, \ t = 0, 1, \ldots \tag{1.2}$$

where m and σ are smooth functions and (ε_t) are assumed to be i.i.d. with mean 0 and variance 1. We briefly illustrate the autoregression bootstrap and the regression bootstrap of which the consistency is investigated by Franke et al. (2002a). These bootstrap procedures commonly resample conditionally (given the original sample) i.i.d. ε_t^*, $t = 1, \ldots, n$ from the distribution obtained by residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$. The residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$ are calculated from nonparametric estimates $\hat{m}$ and $\hat{\sigma}$. The autoregression bootstrap for model (1.2) generates X_t^* by

$$X_t^* = \hat{m}(X_{t-1}^*, \ldots, X_{t-p}^*) + \hat{\sigma}(X_{t-1}^*, \ldots, X_{t-p}^*)\varepsilon_t^* \tag{1.3}$$

and the regression bootstrap for (1.2) generates X_t^* by

$$X_t^* = \hat{m}(X_{t-1}, \ldots, X_{t-p}) + \hat{\sigma}(X_{t-1}, \ldots, X_{t-p})\varepsilon_t^* \tag{1.4}$$

The difference between two procedure is whether to use previously generated $X_{t-1}^*, \ldots, X_{t-p}^*$ as an argument of functions $\hat{m}$, $\hat{\sigma}$ when resampling X_t^* at time t. Note that the bootstrap process from (1.3) has the same structure as the original process (1.2), whereas the bootstrap process from (1.4) does not. Nevertheless, Franke et al. (2002a) have proven the consistency of (1.4) as well as that of (1.3). The local bootstrap proposed by Paparoditis and Politis (2000) also generates the bootstrap samples whose structure is similar to that of (1.4). All these works do not allow discrete variables. In most research on the bootstrap in nonparametric models, they have studied models allowing only continuous variables. We propose the autoregression bootstrap and the regression bootstrap for dynamic discrete response models. The main contribution of this thesis is to propose bootstrap procedures in models that not only admit discrete variables but also allow the lags of the response as regressors.

Bandwidth selection is important in kernel-based approaches. The MSE-optimal bandwidth causes a non-negligible bias, which interferes with valid inference. To tackle this problem, one usually adopts either undersmoothing to reduce the effect of bias or explicit bias correction. Implementation of undersmoothing is simpler than explicit bias correction but it is difficult to propose an optimal bandwidth selection method in practice. For explicit bias correction, practitioners may have to select other bandwidths for the bias corrector in addition to that of the initial estimator. If the bias is complex, oversmoothing is more cumbersome to apply. For detailed literature reviews on undersmoothing and explicit bias correction, see Hall and Horowitz (2013). Hall and Horowitz (2013) proposed a new approach for the construction of bootstrap confidence bands which does not require undersmoothing or explicit bias correction. Comparative studies between undersmoothing and explicit bias correction have been carried out in nonparametric models with i.i.d. observations. According to Hall (1991), Hall (1992) and Neumann (1997), undersmoothing outperforms explicit bias correction for pointwise confidence interval of density and regression function. Recently, Calonico et al. (2018) proposed a robust bias correction and compared it with undersmoothing. However, there have been few studies, up to our knowledge, on models allowing both discrete variables and dependent observations. In this thesis, we focus on undersmoothing. For numerical studies, we use bandwidth which is reduced by a certain amount from the rule-of-thumb bandwidth. Selection of data-driven optimal bandwidth is an interesting topic of future research.

This thesis is organized as follows. In chapter 2, we introduce our models, estimation methodology and its theoretical properties. We show the uniform consistency, expansion and asymptotic normality under suitable conditions. In chapter 3, we describe our bootstrap procedures: the autoregression bootstrap and the regression bootstrap. We also illustrate the consistency of these bootstrap procedures. Chapter 4 contains the results of simulation studies and real data analysis. We consider dynamic binary probit models for binary response data and beta-binomial models as an example of multiparameter models. We illustrate the finite sample performance of our bootstrap procedures through bootstrap confidence intervals. As an application of our methodology, we analyze recession data. We apply dynamic binary probit models and show bootstrap confidence intervals for some points. The proof of theorems and lemmas are presented in the appendix.

Chapter 2

Local Likelihood Estimation

2.1 Models and Estimation

Suppose we observe $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$, $1 \le i \le n$ from a stationary process $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : -\infty < i < \infty\}$. The response variable Y^i is of discrete type. $\mathbf{X}^i = (X_1^i, \ldots, X_d^i)^\top$ is the vector of d-dimensional continuous covariates. $\mathbf{Z}^i$ is the vector of discrete covariates and its components may include d_ℓ lagged values of Y^i. For example, $\mathbf{Z}^i = (Y^{i-d_l}, \ldots, Y^{i-1})$. We employ a local likelihood model

$$Y^i | \mathbf{X}^i = \mathbf{x}, \mathbf{Z}^i = \mathbf{z} \sim f(\cdot, \boldsymbol{\theta}(\mathbf{x}, \mathbf{z}))$$

where f is a pre-specified conditional mass function of Y^i given $\mathbf{X}^i = \mathbf{x}$, $\mathbf{Z}^i = \mathbf{z}$ and $\boldsymbol{\theta}(\mathbf{x}, \mathbf{z}) = (\theta_1, \ldots, \theta_\nu)^\top(\mathbf{x}, \mathbf{z})$ is a unknown $\mathbb{R}^\nu$-valued function. In this section, we give methods for estimating $\boldsymbol{\theta}$ and its first partial derivatives with respect to $\mathbf{x}$.

To obtain estimators of $\boldsymbol{\theta}$, one may consider

$$L_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(Y^i, \boldsymbol{\theta}(\mathbf{X}^i, \mathbf{Z}^i)). \tag{2.1}$$

which is the log-likelihood function of $\boldsymbol{\theta}$ when $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$, $1 \le i \le n$ are i.i.d. However, direct maximization of L_n in an infinite dimensional function space is intractable based on a finite sample. To tackle this problem, we adopt the local linear approach. We can extend to the local polynomial estimation of general order, but we focus on the local linear estimation due to notational simplicity. Let $(\mathbf{x}, \mathbf{z})$ be a point of interest at which we want to estimate the value of $\boldsymbol{\theta}$. For the local linear estimation, we consider the following linear approximation (in the direction of $\mathbf{x}$) of $\boldsymbol{\theta}(\mathbf{u}, \mathbf{v})$ in a neighborhood of $(\mathbf{x}, \mathbf{z})$:

$$\boldsymbol{\theta}(\mathbf{u}, \mathbf{v}) \cong \boldsymbol{\theta}(\mathbf{x}, \mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})(\mathbf{u} - \mathbf{x}) \tag{2.2}$$

where $\boldsymbol{\Theta}(\mathbf{x}, \mathbf{z}) \in \mathbb{R}^{\nu \times d}$ is the matrix of the first partial derivatives of $\boldsymbol{\theta}$ with respect to $\mathbf{x}$, that is, $\Theta_{rj}(\mathbf{u}, \mathbf{v}) = \partial \theta_r(\mathbf{u}, \mathbf{v}) / \partial u_j$ for $\mathbf{u} = (u_1, \ldots, u_d)^\top$. Note that the above approximation is constant in the direction of the discrete covariates $\mathbf{z}$.

Let K be a base line kernel function and define $K_h(u, v) = \frac{1}{h} K(\frac{u-v}{h})$ for $u, v \in \mathbb{R}$ and $h > 0$. We use a product kernel $w_{\mathrm{c}}^i(\mathbf{x}) \times w_{\mathrm{d}}^i(\mathbf{z})$ which is the weight of $(\mathbf{X}^i, \mathbf{Z}^i)$ around $(\mathbf{x}, \mathbf{z})$ where

$$w_{\mathrm{c}}^i(\mathbf{x}) = \prod_{j=1}^{d} K_{h_j}(x_j, X_j^i), \quad w_{\mathrm{d}}^i(\mathbf{z}) = \prod_{j} \lambda_j^{I(Z_j^i \neq z_j)} \tag{2.3}$$

and $I(A)$ is an indicator function such that $I(A) = 1$ if A holds, and zero otherwise. $w_{\mathrm{c}}^i(\mathbf{x})$ and $w_{\mathrm{d}}^i(\mathbf{z})$ are the product kernels for continuous and discrete covariates, respectively, with bandwidths $h_j > 0$ and $0 \le \lambda_j \le 1$. The discrete kernel $w_{\mathrm{d}}^i(\mathbf{z})$ for the discrete covariates $\mathbf{Z}^i$ was employed by Racine and Li (2004). If we take very large λ_j (close to 1) for all j, then $w_{\mathrm{d}}^i(\mathbf{z})$ is close to 1 for all $1 \le i \le n$ and this would be equivalent to ignoring the presence of the discrete covariates $\mathbf{Z}^i$. On the other hand, if we take very small λ_j (close to 0) for all j, then $w_{\mathrm{d}}^i(\mathbf{z})$ is close to 0 for all $1 \le i \le n$ and this would be equivalent to fitting within observations $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : \mathbf{Z}^i = \mathbf{z},\ 1 \le i \le n\}$ and ignoring observations $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : \mathbf{Z}^i \ne \mathbf{z},\ 1 \le i \le n\}$.

To estimate $\boldsymbol{\theta}(\mathbf{x}, \mathbf{z})$ and $\boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})$, utilizing approximation (2.2) and kernels (2.3), we consider

$$L_n(\boldsymbol{\beta}, \mathbf{B}; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \times \log f\left(Y^i, \boldsymbol{\beta} + \mathbf{B}\left(\mathbf{X}^i - \mathbf{x}\right)\right)$$

for $\boldsymbol{\beta} \in \mathbb{R}^{\nu}$ and $\mathbf{B} \in \mathbb{R}^{\nu \times d}$ instead of (2.1). Our local linear estimator $\left(\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}), \hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z})\right)$ can be obtained by maximizing $L_n(\boldsymbol{\beta}, \mathbf{B}; \mathbf{x}, \mathbf{z})$ with respect to $\boldsymbol{\beta}$ and $\mathbf{B}$.

Our models include nonparametric binary choice model which is formulated by $Y^i = I(\theta(\mathbf{X}^i, \mathbf{Z}^i) - \varepsilon^i \ge 0)$ where ε^i is independent of $(\mathbf{X}^i, \mathbf{Z}^i)$ and $\varepsilon^i \sim g^{-1}$ for distribution function g^{-1}. Then, $E(Y^i | \mathbf{X}^i = \mathbf{x}, \mathbf{Z}^i = \mathbf{z}) = g^{-1}(\theta(\mathbf{x}, \mathbf{z}))$. This model has been considered by Park et al. (2017) with link function g. In this model, the likelihood is given by

$$\log f(y, u) = y \log\left(\frac{g^{-1}(u)}{1 - g^{-1}(u)}\right) + \log\left(1 - g^{-1}(u)\right).$$

For a link function g, one usually employ the logit function $g(t) = \log(t/(1-t))$

or probit link $g(t) = \Phi^{-1}(t)$ where Φ is the distribution function of the standard normal distribution.

The nonparametric beta-binomial model is also an example of our models. The probability mass function of the beta-binomial distribution with parameters (N, α, β) is given by

$$\begin{pmatrix} N \\ y \end{pmatrix} \frac{B(y+\alpha, N-y+\beta)}{B(\alpha, \beta)} I(y \in \{0, \dots, N\})$$

where B is the beta function. For simplicity, suppose $Y^i \in \{0, \dots, N\}$. Following Aerts and Claeskens (1997)'s formulation, we employ two link functions $g_1(t) = \log(t/(1-t))$ and $g_2(t) = \log\left(\frac{1+t}{1-t}\right)$. Let $\phi(\mathbf{x}, \mathbf{z}) = \frac{g_2^{-1}(\theta_2(\mathbf{x},\mathbf{z}))}{1-g_2^{-1}(\theta_2(\mathbf{x},\mathbf{z}))}$. In this model, $Y^i|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}$ follows the beta-binomial distribution with parameters $\left(N, g_1^{-1}(\theta_1(\mathbf{x}, \mathbf{z}))/\phi(\mathbf{x}, \mathbf{z}), \left(1 - g_1^{-1}(\theta_1(\mathbf{x}, \mathbf{z}))\right)/\phi(\mathbf{x}, \mathbf{z})\right)$ and the likelihood is given by

$$\begin{aligned}
&\log f(y, (u_1, u_2)^\top) \\
&= \log \begin{pmatrix} N \\ y \end{pmatrix} + \sum_{r=0}^{y-1} \log\left(g_1^{-1}(u_1) + \frac{r g_2^{-1}(u_2)}{1 - g_2^{-1}(u_2)}\right) \\
&\quad + \sum_{r=0}^{N-y-1} \log\left(1 - g_1^{-1}(u_1) + \frac{r g_2^{-1}(u_2)}{1 - g_2^{-1}(u_2)}\right) - \sum_{r=0}^{N-1} \log\left(1 + \frac{r g_2^{-1}(u_2)}{1 - g_2^{-1}(u_2)}\right)
\end{aligned}$$

2.2 Theoretical Properties

In this section, we describe some theoretical properties of the estimator. For this, we introduce some notations. α-mixing coefficient for stationary process $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : -\infty < i < \infty\}$ is defined by

$$\alpha(j) = \sup_{\substack{A \in \mathcal{F}_{-\infty}^0 \\ B \in \mathcal{F}_j^\infty}} |P(A \cap B) - P(A)P(B)|, \ j \geq 1.$$

where $\mathcal{F}_\ell^k$ denotes the σ-field genereted by $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : \ell \leq i \leq k\}$.

For $\boldsymbol{\beta} \in \mathbb{R}^\nu$, let

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\beta}, y) = \frac{\partial \log f(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \in \mathbb{R}^\nu,$$
$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}, y) = \frac{\partial^2 \log f(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \in \mathbb{R}^{\nu \times \nu}.$$

Let supp($\mathbf{U}$) denote the support of a random vector $\mathbf{U}$ and $(\mathbf{X}, \mathbf{Z}, Y)$ be the triple such that $(\mathbf{X}, \mathbf{Z}, Y) \stackrel{\mathrm{d}}{=} (\mathbf{X}^i, \mathbf{Z}^i, Y^i)$. To obtain the theoretical properties, we use the following conditions.

(C1) $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : -\infty < i < \infty\}$ is a stationary process.

(C2) The marginal density of $\mathbf{X}$ has support $[0, 1]^d$.

(C3) The joint density $p_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ of $(\mathbf{X}, \mathbf{Z})$ is continuously differentiable in $\mathbf{x}$ for each $\mathbf{z}$ and is bounded away from zero on its support.

(C4) Both the conditional density of $(\mathbf{X}^i, \mathbf{Z}^i)$ given Y^i and the conditional density of $(\mathbf{X}^i, \mathbf{Z}^i, \mathbf{X}^{i+l}, \mathbf{Z}^{i+l})$ given (Y^i, Y^{i+l}) exist and are bounded.

(C5) The base kernel K is a continuously differentiable symmetric density function supported on $[-1, 1]$.

(C6) $h_{\max} = \max_{1 \le j \le d} h_j \to 0$ and $\lambda_{\max} = \max_j \lambda_j \to 0$ as $n \to \infty$.

(C7) $\tau_n = (nh_1 \times \cdots \times h_d / \log n)^{1/2} \to \infty$ as $n \to \infty$.

(C8) $n\upsilon_n^{\nu(d+1)} (\tau_n h_1 \times \cdots \times h_d)^{-1/2} \alpha(\tau_n) \to 0$ as $n \to \infty$ where $\upsilon_n = \tau_n(h_1 \times \cdots \times h_d \times \min_{1 \le j \le d} h_j)^{-1}$.

(C9) $\sum_{n=1}^{\infty} n^b \alpha(n)^{1-a/2} < \infty$ for some $a > 2$, $b > 1 - a/2$.

(C10) For each $(\mathbf{x}, \mathbf{z}) \in \text{supp}(\mathbf{X}, \mathbf{Z})$, $Y | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim f(\cdot, \boldsymbol{\theta}(\mathbf{x}, \mathbf{z}))$.

(C11) For each $\mathbf{z} \in \text{supp}(\mathbf{Z})$, $\boldsymbol{\theta}(\cdot, \mathbf{z})$ is three times continuously differentiable on $\text{supp}(\mathbf{X})$.

(C12) For each $y \in \text{supp}(Y)$, $\log f(y, \cdot)$ is three times continuously differentiable.

(C13) $\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) = -E\left[\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{Z}), Y) | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}\right]$ is continuously differentiable in $\mathbf{x}$ for each $\mathbf{z}$ and its smallest eigen values are bounded away from zero on $\text{supp}(\mathbf{X}, \mathbf{Z})$.

Note that $|\text{supp}(Y)|$, $|\text{supp}(\mathbf{Z})| < \infty$ by (C3) which is a standard condition in nonparametric models. By $|\text{supp}(Y)| < \infty$ and (C12), we obtain versions of Bartlett's identities:

$$E\left[\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{Z}), Y) | \mathbf{X}, \mathbf{Z}\right] = \sum_y \dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{Z}), y) f(y, \boldsymbol{\theta}(\mathbf{X}, \mathbf{Z})) = 0$$

$$
\begin{aligned}
-E\left[\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),Y)|\mathbf{X},\mathbf{Z}\right] &= -\sum_y \ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),y)f(y,\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}))\\
&= \sum_y \dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),y)\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),y)^\top f(y,\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}))\\
&= E\left[\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),Y)\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X},\mathbf{Z}),Y)^\top|\mathbf{X},\mathbf{Z}\right]
\end{aligned}
$$

(C8) and (C9) are conditions for the decaying rate of the α-mixing coefficient. These conditions are modified from the conditions in Masry (1996) which derived the uniform consistency of the local polynomial estimator with α-mixing processes. Decaying rate of the α-mixing coefficient satisfying (C8) and (C9) is related to the magnitude of the bandwidths. For example,

$$
\alpha(n) = O\left(n^{-[2\nu(d+1)(d+3)+3(d+2)]/4}\right)
$$

is enough to take bandwidths $h_j \sim n^{-1/(d+4)}$, $\lambda_{j'} \sim n^{-2/(d+4)}$ (MSE-optimal bandwidths for θ_j from Lemma 2.3) or $h_j \sim n^{-1/(d+4)}(\log n)^{-\varepsilon}$, $\lambda_{j'} \sim n^{-2/(d+4)}(\log n)^{-\varepsilon}$ for some $\varepsilon > 0$ (smaller bandwidths than MSE-optimal bandwidths for undersmoothing).

Let $\iota(\boldsymbol{\beta},\mathbf{B}) \in \mathbb{R}^{\nu(d+1)}$ denote a vector obtained by concatenating the entries of $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_\nu)^\top \in \mathbb{R}^\nu$ and $\mathbf{B} \in \mathbb{R}^{\nu\times d}$. It is defined by $\iota(\boldsymbol{\beta},\mathbf{B}) = (\beta_1,\mathbf{B}_1,\ldots,\beta_\nu,\mathbf{B}_\nu)^\top$ where d-dimensional row vector $\mathbf{B}_j$ is a j-th row of $\mathbf{B}$, $j = 1,\ldots,\nu$. The following theorem is on the existence of at least one solution to the local log likelihood equations and its uniform consistency.

Theorem 2.1. *Suppose conditions (C1) $\sim$ (C13). Then, there exists a solution* $\left(\hat{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}),\hat{\boldsymbol{\Theta}}(\mathbf{x},\mathbf{z})\right)$ *to the equation* $\partial L_n(\boldsymbol{\beta},\mathbf{B};\mathbf{x},\mathbf{z})/\partial(\iota(\boldsymbol{\beta},\mathbf{B})) = \mathbf{0}$ *for all*

$(\mathbf{x},\mathbf{z}) \in supp(\mathbf{X},\mathbf{Z})$ *with probability tending to one. Moreover, we have*

$$\sup_{(\mathbf{x},\mathbf{z})\in supp(\mathbf{X},\mathbf{Z})} \left|\hat{\theta}_r(\mathbf{x},\mathbf{z}) - \theta_r(\mathbf{x},\mathbf{z})\right| = O_P\left(\tau_n^{-1} + h_{\max}^2 + \lambda_{max}\right)$$

$$\sup_{(\mathbf{x},\mathbf{z})\in supp(\mathbf{X},\mathbf{Z})} \left|h_j\left(\hat{\Theta}_{rj}(\mathbf{x},\mathbf{z}) - \Theta_{rj}(\mathbf{x},\mathbf{z})\right)\right| = O_P\left(\tau_n^{-1} + h_{\max}^2 + \lambda_{max}\right)$$

for all $r = 1,\ldots,\nu$ *and* $j = 1,\ldots,d$.

We next describe the expansion of our estimator. For this, let

$$\mathbf{W}^i = \left(1, (X_1^i - x_1)/h_1, \ldots, (X_d^i - x_d)/h_d\right)^\top \in \mathbb{R}^{d+1}$$

$$\mathbf{H} = \mathrm{diag}(h_1,\ldots,h_d) \in \mathbb{R}^{d\times d}.$$

For two matrices $\mathbf{A} = (a_{jk}) \in \mathbb{R}^{m_1\times m_2}$ and $\mathbf{B}$, define the Kronecker product $\otimes$ by

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,m_2}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m_1,1}\mathbf{B} & \cdots & a_{m_1,m_2}\mathbf{B} \end{pmatrix}$$

Define

$$\mathbf{G}(\mathbf{x},\mathbf{z}) = -E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{x},\mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x},\mathbf{z})(\mathbf{X}^i - \mathbf{x}), Y^i\right)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\right]$$

$$\hat{\mathbf{F}}(\mathbf{x},\mathbf{z}) = n^{-1}\sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{x},\mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x},\mathbf{z})(\mathbf{X}^i - \mathbf{x}), Y^i\right)\otimes\mathbf{W}^i$$

Lemma 2.2. *Under the conditions of Theorem 2.1, we have*

$$\begin{aligned}&\iota\left(\hat{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}) - \boldsymbol{\theta}(\mathbf{x},\mathbf{z}), \left[\hat{\boldsymbol{\Theta}}(\mathbf{x},\mathbf{z}) - \boldsymbol{\Theta}(\mathbf{x},\mathbf{z})\right]\mathbf{H}\right)\\ &= \left(\mathbf{G}(\mathbf{x},\mathbf{z})^{-1} + o_P(1)\right)\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})\end{aligned}$$

for each $(\mathbf{x},\mathbf{z}) \in supp(\mathbf{X},\mathbf{Z})$.

We apply Lemma 2.2 to construct the bootstrap estimator in chapter 3 as well as to derive the asymptotic distribution of the estimator. We first give the asymptotic distribution of $\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})$. For this, let $(L)_{\mathbf{g}}(\mathbf{v},\mathbf{w}) = \frac{1}{g_1\cdots g_d} L\left(\frac{v_1-w_1}{g_1}\right)\cdots L\left(\frac{v_d-w_d}{g_d}\right)$ for $\mathbf{g}=(g_j)$, $\mathbf{v}=(v_j)$, $\mathbf{w}=(w_j) \in \mathbb{R}^d$ and a univariate function L. Define

$$\mathbf{M}(\mathbf{x}) = \int \mathbf{u}_n \mathbf{u}_n^\top \left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})\mathrm{d}\mathbf{u} \in \mathbb{R}^{(d+1)\times(d+1)},$$
$$\mathbf{N}(\mathbf{x}) = \int \mathbf{u}_n \mathbf{u}_n^\top (K)_{\mathbf{h}}(\mathbf{x},\mathbf{u})\mathrm{d}\mathbf{u} \in \mathbb{R}^{(d+1)\times(d+1)}.$$

where $\mathbf{u}_n = \left(1, \frac{u_1-x_1}{h_1}, \ldots, \frac{u_d-x_d}{h_d}\right)^\top$ and $\mathbf{h}=(h_j)$. Note that for interior point $\mathbf{x}$, both $\mathbf{M}(\mathbf{x})$ and $\mathbf{N}(\mathbf{x})$ are diagonal matrices. In this case, $(1,1)$ entry of $\mathbf{M}(\mathbf{x})$ is $\left(\int K^2\right)^d$ and (j,j) entry of that is $\int u^2 K^2(u)\mathrm{d}u \times \left(\int K^2\right)^{d-1}$ for $2 \le j \le d$. $(1,1)$ entry of $\mathbf{N}(\mathbf{x})$ is 1 and (j,j) entry of that is $\int u^2 K(u)\mathrm{d}u$ for $2 \le j \le d$ in the same case. Let $\mathbf{I}_k \in \mathbb{R}^{k\times k}$ be the identity matrix for $k \in \mathbb{N}$. Since we adopt undersmoothing in chapter 3, we do not focus on the bias part of the estimators. In Lemma 2.3, we just give the magnitude of $E\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})$.

Lemma 2.3. *Under the conditions of Theorem 2.1, we have for all* $(\mathbf{x},\mathbf{z}) \in supp(\mathbf{X},\mathbf{Z})$

$$(nh_1 \times \cdots \times h_d)^{1/2} p_{\mathbf{X},\mathbf{Z}}^{-1/2} \left(\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}) \otimes \mathbf{M}(\mathbf{x})\right)^{-1/2} \left(\hat{\mathbf{F}}(\mathbf{x},\mathbf{z}) - E\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})\right) \xrightarrow{\mathrm{d}} N(0, \mathbf{I}_{\nu(d+1)})$$

and $E\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z}) = O\left(h_{\max}^2 + \lambda_{\max}\right)$ *for all* $\mathbf{x}$, $\mathbf{z}$. *If* $\mathbf{x}$ *is the interior point of* *supp*$(\mathbf{X})$, *the* j*-th component of* $E\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z})$ *is of the magnitude* $O\left(h_{\max}^3 + \lambda_{\max}\right)$ *for* $j \in \{1, \ldots, \nu(d+1)\} \setminus \{1, d+2, \ldots, (\nu-1)(d+1)+1\}$.

We have

$$\mathbf{G}(\mathbf{x}, \mathbf{z}) = -p_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})\left[\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) \otimes \mathbf{N}(\mathbf{x})\right] + o(1) \tag{2.4}$$

Combining Lemma 2.2, Lemma 2.3 and (2.4), we get the asymptotic distribution of $\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})$ and $\hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z})$ under the conditions of Theorem 2.1. We summarize the asymptotic variance and the asymptotic bias of $\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})$ and $\hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z})$. In the case of $\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})$, for each $(\mathbf{x}, \mathbf{z}) \in \mathrm{supp}(\mathbf{X}, \mathbf{Z})$, the asymptotic variance and the asymptotic bias is of order $O\left((nh_1 \times \cdots \times h_d)^{-1}\right)$ and $O(h_{\max}^2 + \lambda_{\max})$, respectively. In the case of $\hat{\Theta}_{rj}(\mathbf{x}, \mathbf{z})$, the asymptotic variance is of the magnitude $O\left((nh_1 \times \cdots \times h_d)^{-1} h_j^{-2}\right)$ for each $(\mathbf{x}, \mathbf{z}) \in \mathrm{supp}(\mathbf{X}, \mathbf{Z})$. The asymptotic bias of $\hat{\Theta}_{rj}(\mathbf{x}, \mathbf{z})$ is of order $O\left(h_j^{-1}(h_{\max}^2 + \lambda_{\max})\right)$ if $\mathbf{x}$ is the boundary point of supp$(\mathbf{X})$, and of order $O\left(h_j^{-1}(h_{\max}^3 + \lambda_{\max})\right)$ otherwise.

Chapter 3

Bootstrap

3.1 Methodology

In this section, we provide our bootstrap procedures. Franke et al. (2002a) investigated several bootstrap procedures in nonlinear time series model. In especially, they have shown the properties of the autoregression bootstrap and the regression bootstrap procedures. These approaches are residual-based resampling. They first estimate the unknown functions using kernel smoothing and then obtain residuals. From conditional distribution constructed from these residuals, they resample a conditionally (given the original sample) i.i.d. variables. Utilizing resampled residuals, a bootstrap process is generated whose values at time t depend on some bootstrap samples in previous time in the autoregression bootstrap. On the other hand, in the regression bootstrap, a bootstrap process is generated whose values at time t only depend on the original sample and do not depend on the other values from the bootstrap sample. In this procedure, the bootstrap samples are conditionally independent given the original sample. Although the regression bootstrap seems not to work since bootstrap samples generated from this procedure do not have the same structure of the original process, Franke et al. (2002a) has proven that the regression bootstrap procedures as well as the autoregression bootstrap consistently approximate the laws of the estimators of interest. In fact, according to Franke et al. (2002a), the proof for the regression bootstrap is simpler than that of the autoregression bootstrap. Since the stochastic structure of the autoregression bootstrap is random (not fixed) and therefore, it is difficult to apply the techniques of mixing processes.

Since the response variable in our model is of discrete, we can not directly apply the residual-based resampling as in Franke et al. (2002a). Instead, we generate the response variable from the likelihood model and estimated values. We also provide two bootstrap procedures in dynamic discrete response models. To introduce our approaches, write $\mathbf{Z}^i = (\mathbf{Z}_1^i, \mathbf{Z}_2^i)$ where $\mathbf{Z}_1^i$ is a vector of exogenous discrete variables and $\mathbf{Z}_2^i = (Y^{i-d_l}, \ldots, Y^{i-1})$ is a vector of d_l lags of the dependent variable.

In the first bootstrap procedure which we call the autoregression bootstrap, we generate a bootstrap sample $(\mathbf{X}^i, \mathbf{Z}^{i,*}, Y^*)$, $1 \leq i \leq n$ by

(AB1) Initialize $\left(Y^{1-d_l,*}, \ldots, Y^{0,*}\right) = \mathbf{Z}_2^1$.

(AB2) For $i = 1, \ldots, n$, sequentially put $\mathbf{Z}_2^{i,*} = (Y^{i-d_l,*}, \ldots, Y^{i-1,*})$ and draw

$Y^{i,*}$ from $f(\cdot, \hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i,*})$.

where $\mathbf{Z}^{i,*} = (\mathbf{Z}_1^i, \mathbf{Z}_2^{i,*})$ and $\hat{\boldsymbol{\theta}}$ is the local linear estimator of $\boldsymbol{\theta}$ proposed in chapter 2. In this resampling scheme, a bootstrap sample mimics the structure of the original sample as in the autoregression bootstrap by Franke et al. (2002a). When we generate an autoregression bootstrap sample in practice, we do not resample values of exogenous variables $\mathbf{X}^i$ and $\mathbf{Z}_1^i$, but use the values in the original sample. We recursively update $Y^{i,*}$ using its lagged values $\mathbf{Z}_2^{i,*}$. Note that there exist $\hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i,*})$ for all $1 \le i \le n$ with probability tending to one by Theorem 2.1.

In the second bootstrap procedure (the regression bootstrap), we generate a bootstrap sample $(\mathbf{X}^i, \mathbf{Z}^i, Y^*)$, $1 \le i \le n$ from

(RB) conditionally independent (given $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : i = 1, \ldots, n\}$)

$Y^{i,*}$, $i = 1, \ldots, n$ are drawn from $f(\cdot, \hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^i))$.

Note that $Y^{i,*}$ does not depend on $(Y^{i-d_l,*}, \ldots, Y^{i-1,*})$ in the regression bootstrap.

The goal of the bootstrap is to get the approximation of the distribution of the estimator. For this, we construct the bootstrap estimator using a bootstrap sample. Let $(\mathbf{x}, \mathbf{z})$ be fixed. One may consider an estimator maximizing the bootstrap analogue of L_n:

$$L_n^*(\boldsymbol{\beta}, \mathbf{B}; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^{i,*}(\mathbf{z}) \times \log f\left(Y^{i,*}, \boldsymbol{\beta} + \mathbf{B}\left(\mathbf{X}^i - \mathbf{x}\right)\right) \text{ for (AB)},$$

$$L_n^*(\boldsymbol{\beta}, \mathbf{B}; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \times \log f\left(Y^{i,*}, \boldsymbol{\beta} + \mathbf{B}\left(\mathbf{X}^i - \mathbf{x}\right)\right) \text{ for (RB)}$$

where $w_{\mathrm{d}}^{i,*}(\mathbf{z}) = \prod_j \lambda_j^{I(Z_j^{i,*} \neq z_j)}$. However, this method requires too many iterative calculations. To avoid this computational burden, we propose a one-step bootstrap estimator. Motivated from Lemma 2.2, define the bootstrap analogue of $\mathbf{G}(\mathbf{x}, \mathbf{z})$ and $\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z})$ by

$$\hat{\mathbf{G}}(\mathbf{x}, \mathbf{z}) = -n^{-1} \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) + \hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z})\left(\mathbf{X}^i - \mathbf{x}\right), Y^i\right) \otimes \mathbf{W}^i (\mathbf{W}^i)^\top$$

$$\hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) = \begin{cases} n^{-1} \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^{i,*}(\mathbf{z}) \dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i,*}), Y^{i,*}\right) \otimes \mathbf{W}^i \text{ for (AB)} \\ n^{-1} \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i}), Y^{i,*}\right) \otimes \mathbf{W}^i \text{ for (RB)} \end{cases}$$

We define $\hat{\boldsymbol{\theta}}^*(\mathbf{x}, \mathbf{z})$ and $\hat{\boldsymbol{\Theta}}^*(\mathbf{x}, \mathbf{z})$ by

$$\iota\left(\hat{\boldsymbol{\theta}}^*(\mathbf{x}, \mathbf{z}) - \hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}), \left[\hat{\boldsymbol{\Theta}}^*(\mathbf{x}, \mathbf{z}) - \hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z})\right] \mathbf{H}\right) = \hat{\mathbf{G}}(\mathbf{x}, \mathbf{z})^{-1} \hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z})$$

, which is a bootstrap estimator of $\boldsymbol{\theta}(\mathbf{x}, \mathbf{z})$ and $\boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})$.

3.2 Theoretical Properties

In this section, we discuss the consistency of the bootstrap procedures. We adopt undersmoothing to avoid bias estimation. For undersmoothing, we use condition

(U1) $(nh_1 \times \cdots \times h_d)^{1/2} (h_{\max}^2 + \lambda_{\max}) \to 0$ as $n \to \infty$,

(U2) $(nh_1 \times \cdots \times h_d)^{1/2} (h_{\max}^3 + \lambda_{\max}) \to 0$ as $n \to \infty$.

Let d_{K} denote the Kolmogorov distance(the supremum metric between two corresponding distribution functions). Let $\mathcal{L}(W)$ be the distribution of W and $\mathcal{L}^*(W)$ be the conditional distribution of W given $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : i = 1, \ldots, n\}$ for a random variable W. For each bootstrap procedure, define

$$\begin{aligned}
\mathcal{L}_r(\mathbf{x}, \mathbf{z}) &= \mathcal{L}\left((nh_1 \times \cdots \times h_d)^{1/2} \left(\hat{\theta}_r(\mathbf{x}, \mathbf{z}) - \theta_r(\mathbf{x}, \mathbf{z})\right)\right), \\
\mathcal{L}_r^*(\mathbf{x}, \mathbf{z}) &= \mathcal{L}^*\left((nh_1 \times \cdots \times h_d)^{1/2} \left(\hat{\theta}_r^*(\mathbf{x}, \mathbf{z}) - \hat{\theta}_r(\mathbf{x}, \mathbf{z})\right)\right), \\
\mathcal{L}_{rj}(\mathbf{x}, \mathbf{z}) &= \mathcal{L}\left((nh_1 \times \cdots \times h_d)^{1/2} h_j \left(\hat{\Theta}_{rj}(\mathbf{x}, \mathbf{z}) - \Theta_{rj}(\mathbf{x}, \mathbf{z})\right)\right), \\
\mathcal{L}_{rj}^*(\mathbf{x}, \mathbf{z}) &= \mathcal{L}^*\left((nh_1 \times \cdots \times h_d)^{1/2} h_j \left(\hat{\Theta}_{rj}^*(\mathbf{x}, \mathbf{z}) - \hat{\Theta}_{rj}(\mathbf{x}, \mathbf{z})\right)\right)
\end{aligned}$$

for $1 \le r \le \nu$ and $1 \le j \le d$. We first consider the consistency of the regression bootstrap.

Theorem 3.1. *In the case of the regression bootstrap, if the conditions of*

Theorem 2.1 and (U1) are satisfied, then for each $(\mathbf{x}, \mathbf{z}) \in supp(\mathbf{X}, \mathbf{Z})$

$$\begin{aligned} d_{\mathrm{K}}\left(\mathcal{L}_r(\mathbf{x}, \mathbf{z}), \mathcal{L}_r^*(\mathbf{x}, \mathbf{z})\right) &= o_P(1), \\ d_{\mathrm{K}}\left(\mathcal{L}_{rj}(\mathbf{x}, \mathbf{z}), \mathcal{L}_{rj}^*(\mathbf{x}, \mathbf{z})\right) &= o_P(1) \end{aligned} \tag{3.1}$$

for all $r = 1, \ldots, \nu$ *and* $j = 1, \ldots, d$.

For each $\mathbf{x}$ in the interior of supp($\mathbf{X}$) and $\mathbf{z} \in$ supp($\mathbf{Z}$), the bias of $h_j \hat{\Theta}_{rj}(\mathbf{x}, \mathbf{z})$ is of the magnitude $O(h_{\max}^3 + \lambda_{\max})$ as we have stated below Lemma 2.3. Therefore, in this case, we have (3.1) under (U2) which is a weaker condition than (U1).

In the case of the autoregression bootstrap, we need additional conditions:

(C1*) For each $\mathbf{x}^1, \ldots, \mathbf{x}^i, \mathbf{z}^1, \ldots, \mathbf{z}^i$ and y,

$$P\left(Y^i = y | \mathbf{X}^i = \mathbf{x}^i, \ldots, \mathbf{X}^1 = \mathbf{x}^1, \mathbf{Z}^i = \mathbf{z}^i, \ldots, \mathbf{Z}^1 = \mathbf{z}^1\right) = f(y, \boldsymbol{\theta}(\mathbf{x}^i, \mathbf{z}^i)).$$

(C2*) $\mathbf{Z}_2^i$ and $(\mathbf{X}^i, \mathbf{Z}_1^i)$ are conditionally independent given $\mathbf{X}^{i-1}, \ldots, \mathbf{X}^1$ and $\mathbf{Z}_1^{i-1}, \ldots, \mathbf{Z}_1^1$.

(C3*) There exists $C_1 > \frac{d_Y - 2}{d_Y(d_Y - 1)}$ such that $P(Y = y | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \geq C_1$ for all $y \in$ supp(Y), $(\mathbf{x}, \mathbf{z}) \in$ supp($\mathbf{X}, \mathbf{Z}$) where $d_Y = |\mathrm{supp}(Y)|$.

Note that

$$0 < C_1 \leq P(Y = y | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) \leq 1 - (d_Y - 1)C_1 < 1$$

for all $y \in \mathrm{supp}(Y)$, $(\mathbf{x}, \mathbf{z}) \in \mathrm{supp}(\mathbf{X}, \mathbf{Z})$ and

$$\frac{d_Y - 2}{d_Y(d_Y - 1)} < C_1 \leq \frac{1}{d_Y}$$

With these conditions, we are also able to show the consistency of the autoregression bootstrap.

Theorem 3.2. *In the case of (AB), if the conditions of Theorem 2.1, (C1*), (C2*), (C3*) and (U1) are satisfied, then for each* $(\mathbf{x}, \mathbf{z}) \in$ *supp*$(\mathbf{X}, \mathbf{Z})$

$$\begin{aligned} &d_{\mathrm{K}}\left(\mathcal{L}_r(\mathbf{x}, \mathbf{z}), \mathcal{L}_r^*(\mathbf{x}, \mathbf{z})\right) = o_P(1), \\ &d_{\mathrm{K}}\left(\mathcal{L}_{rj}(\mathbf{x}, \mathbf{z}), \mathcal{L}_{rj}^*(\mathbf{x}, \mathbf{z})\right) = o_P(1) \end{aligned} \tag{3.2}$$

for all $r = 1, \ldots, \nu$ *and* $j = 1, \ldots, d$.

As in the case of Theorem 3.1, for each $\mathbf{x}$ in the interior of $\mathrm{supp}(\mathbf{X})$ and $\mathbf{z} \in \mathrm{supp}(\mathbf{Z})$, we get (3.2) if $(U2)$ is satisfied instead of $(U1)$.

Chapter 4

Numerical Studies

4.1 Simulation

In this section, we illustrate the finite sample performance of the proposed bootstrap procedures. We focus on the performance of confidence intervals based on our bootstrap procedures. In the first setting, the model of interest is the dynamic binary probit model (described in section 2.1):

$$Y^i | X^i = x, Z^i = z \sim \mathcal{B}(\Phi(\theta(x, z))) \tag{4.1}$$

for $i = 1, \ldots, n$ where $Z^i = Y^{i-1}$, Φ is a distribution function of standard normal distribution. We generate data consisting of $n = 500$ samples from (4.1) with

$$\theta(x, z) = 0.3 + 0.5 \sin(-1.8x + 2(z - 1))$$

and $X^i \sim \mathcal{U}(-2,2)$. We consider 4 points: $(x,z) \in \{(-1,1),(0,0),(1,1),(1.5,0)\}$. The function and evaluated points are plotted in Figure 4.1. Bandwidth selection is an important problem in nonparametric regression. As conducted by Park et al. (2017), we first put the rule-of-thumb bandwidths $h_1 = 1.06 \times n^{-1/5}\hat{\sigma}_X$ and $\lambda_1 = n^{-2/5}$. For undersmoothing, we consider three bandwidths: $h_1 \times (\log n)^{-\delta}$, $\lambda_1 \times (\log n)^{-\delta}$ for $\delta = 0.1, 0.2, 0.3$ where $\hat{\sigma}_X$ is the sample standard deviation of X. For $n = 500$, these factors are $(\log n)^{-0.1} \cong 0.83$, $(\log n)^{-0.2} \cong 0.69$ and $(\log n)^{-0.3} \cong 0.58$. We compare empirical coverages and average lengths of 95% confidence intervals based on the autoregression bootstrap (AB) and the regression bootstrap (RB). For each bootstrap procedure, a bootstrap confidence interval is calculated from $B = 500$ bootstrap estimates. We carry out $M = 500$ simulation runs, that is, to obtain an empirical coverage and average length we conduct followings:
for each bootstrap procedure, point and bandwidth

1. we generate M replicated datasets with sample size n

2. for each dataset we obtain bootstrap confidence interval from B bootstrap estimates.

3. Finally, we compute empirical coverage and average length from M confidence intervals.

In Table 4.1 and Table 4.2, we illustrate the performance of the confidence intervals based on our bootstrap procedures. We observe that empirical coverages of our bootstrap confidence intervals are close to the nominal value for all 4 points. We also see that average lengths of confidence interval of the regression bootstrap are shorter than those of the autoregression bootstrap, while there is little difference in empirical coverages between two procedures. In further research, it is necessary to conduct experiments for discrete choice models encompassing several exogenous covariates and lagged values of the response variable.

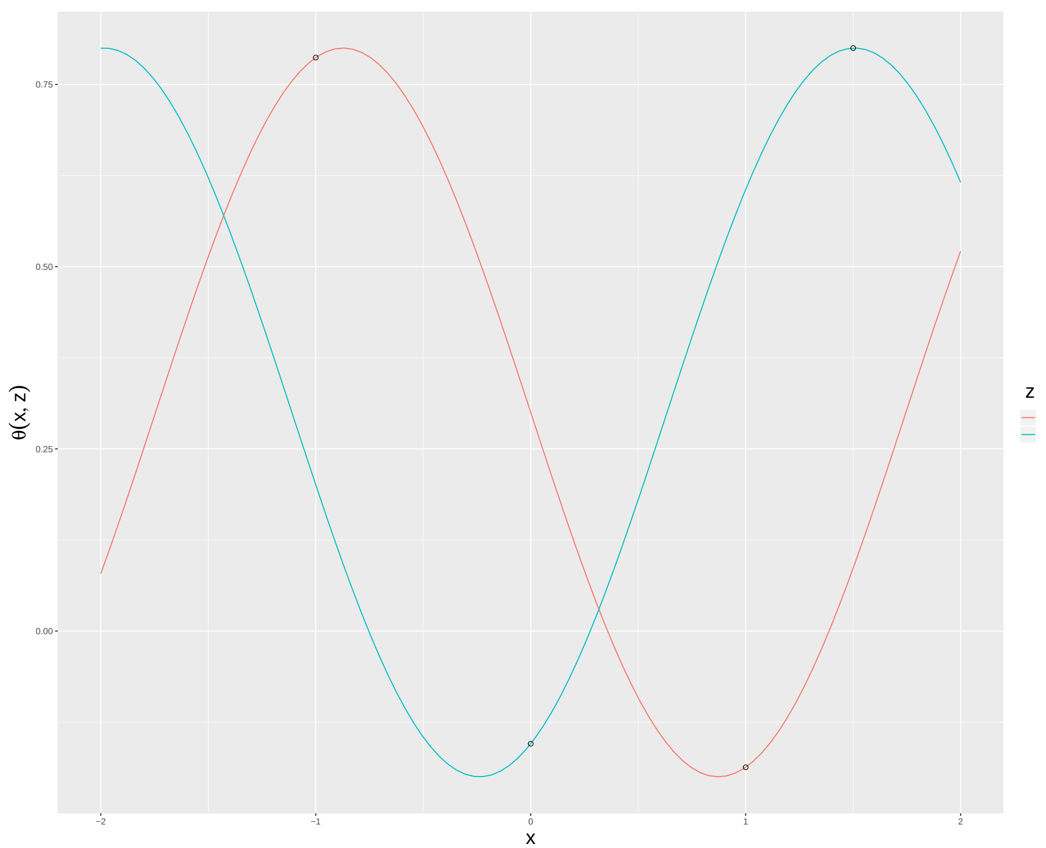

Figure 4.1: The function $\theta(x, z)$ with 4 evaluation points in setting 1.

		Coverage		Length	
(x, z)	δ	AB	RB	AB	RB
(-1,1)	0.1	0.954	0.962	0.919	0.910
	0.2	0.944	0.944	1.043	1.025
	0.3	0.952	0.956	1.221	1.187
(0,0)	0.1	0.950	0.940	1.001	0.959
	0.2	0.936	0.934	1.141	1.085
	0.3	0.942	0.932	1.315	1.246

Table 4.1: Empirical coverages and average lengths of 95% confidence intervals for point $(-1, 1)$ and $(0, 0)$ in binary probit model

		Coverage		Length	
(x, z)	δ	AB	RB	AB	RB
(1,1)	0.1	0.938	0.940	0.802	0.787
	0.2	0.948	0.946	0.890	0.871
	0.3	0.940	0.934	0.993	0.973
(1.5,0)	0.1	0.928	0.920	1.101	1.077
	0.2	0.958	0.930	1.312	1.258
	0.3	0.944	0.936	1.601	1.511

Table 4.2: Empirical coverages and average lengths of 95% bootstrap confidence intervals for point $(1, 1)$ and $(1.5, 0)$ in binary probit model

Next, we consider the beta-binomial model which is also described in section 2.1. We follow the notations in section 2.1. For simplicity, we assume $N = 2$, that is, $Y^i \in \{0, 1, 2\}$ for all $1 \le i \le n$. We generate data from

$$\theta_1(x, z) = -0.5z + 0.3(z - 1)x^2$$
$$\theta_2(x, z) = 0.7 + 0.2\sin(x - 2(z + 0.3))$$

and $X^i \sim \mathcal{U}(-2, 2)$. The others are similar to setting 1. We consider three points: $(x, z) \in \{(-1, 0), (0, 1), (1, 2)\}$. The functions and evaluated points are plotted in Figure 4.2. We put the rule-of-thumb bandwidths $h_1 = 1.06 \times n^{-1/5}\hat{\sigma}_X$, $\lambda_1 = n^{-2/5}$ and consider two bandwidths: $h_1 \times \log(n)^{-\delta}$, $\lambda_1 \times \log(n)^{-\delta}$ for $\delta = 0.1, 0.2$. We put $M = 300$ and $B = n = 500$. Table 4.3 shows the coverage and lengths of 95% confidence intervals based on our bootstrap procedures for θ_1 at three evaluation points. We observe that empirical coverages of our bootstrap confidence intervals are close to the nominal value for all 3 points. In especially, empirical coverages of the autoregression bootstrap are slightly closer to the nominal value than those of the regression bootstrap. As in the case of setting 1, average lengths of confidence intervals of the regression bootstrap are shorter than those of the autoregression bootstrap.

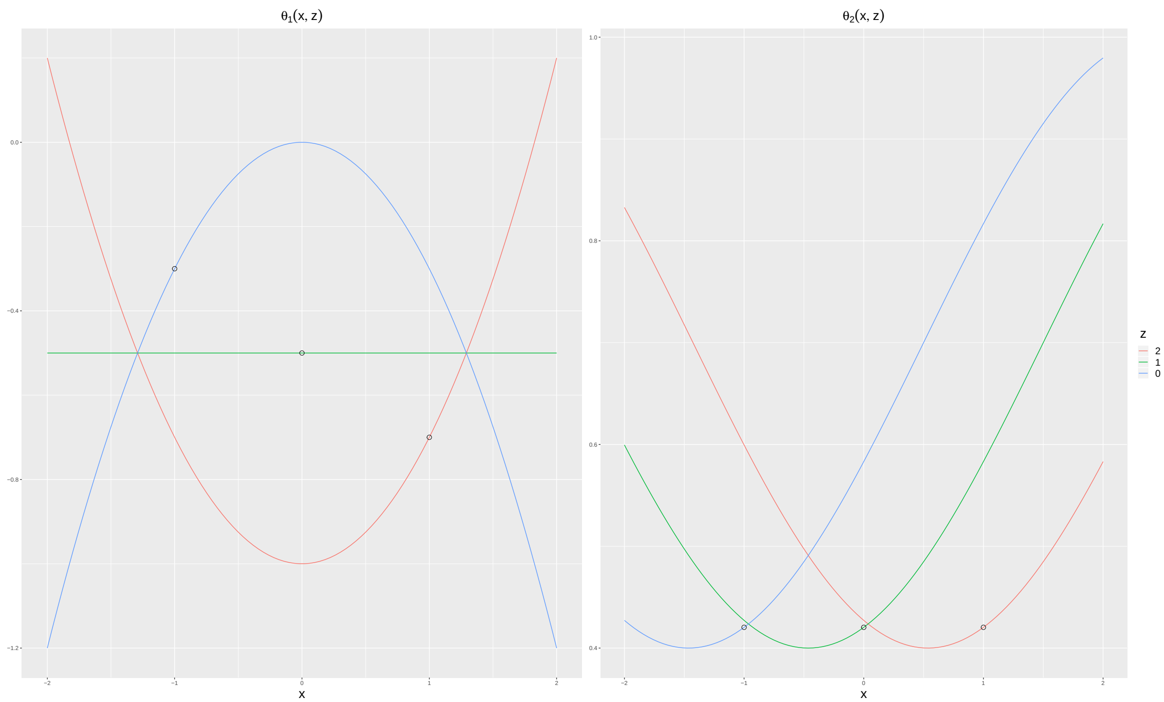

Figure 4.2: Two true functions $\theta_1(x, z)$(left) and $\theta_2(x, z)$(right) with 3 evluation points.

		Coverage		Length	
(x, z)	δ	AB	RB	AB	RB
(-1,0)	0.1	0.950	0.943	1.233	1.192
	0.2	0.920	0.920	1.409	1.357
(0,1)	0.1	0.943	0.943	1.527	1.464
	0.2	0.947	0.940	1.790	1.692
(1,2)	0.1	0.957	0.957	1.918	1.782
	0.2	0.943	0.937	2.400	2.082

Table 4.3: Empirical coverages and average lengths of 95% bootstrap confidence intervals of $\theta_1(x, z)$ in beta-binomial model

4.2 Real Data Analysis

Applying two bootstrap procedures, we analyze data on US recession. Kauppi and Saikkonen (2008) applied parametric dynamic binary probit model to US recession data and Park et al. (2019) did nonparametric dynamic binary probit model developed by Park et al. (2017). We also use the same data sources with these two works and follow the definitions of variables. The response variable $Y^i = 1$ if US economy is considered as in recession in the quarter i and 0 otherwise. The definition of the recession is different in each literature and we follow the definitions of two works. Y^i is constructed from "business cycle peak" and "business cycle trough" data from NBER's Business Cycle Dating Committee (see http://www.nber.org/cycles). We also use a continuous regressor X^i defined by the difference between the 10-year US Treasury bond rate and the 3-month US Treasury bill rate from Federal Reserve Bank. (We extracted these data from https://www.federalreserve.gov/datadownload/Choose.aspx?rel=H15) For details for data, see Park et al. (2019). They also applied one lagged value of the dependent variable and consequently we consider dynamic binary probit model

$$E\left(Y^i | X^i = x, Z^i = z\right) = \Phi(\theta(x, z))$$

where $Z^i = Y^{i-1}$ and Φ is a distribution function of standard normal distribution.

The data consit of $221 (= n)$ quarterly observations (1955:Q4 $\sim$ 2010:Q4) on US recessions(Y^i) and the spread($\mathbf{X}^i$). We illustrate our estimates in Figure

4.3 by time. Figure 4.4 also shows the estimated $\Phi(\hat{\theta}(x,z))$ for each point (x,z). We also give 95% confidence intervals based on the autoregression bootstrap for points (x,z). The range of x is selected from Q1 of $X-0.5$ to Q3 of $X+0.5$.

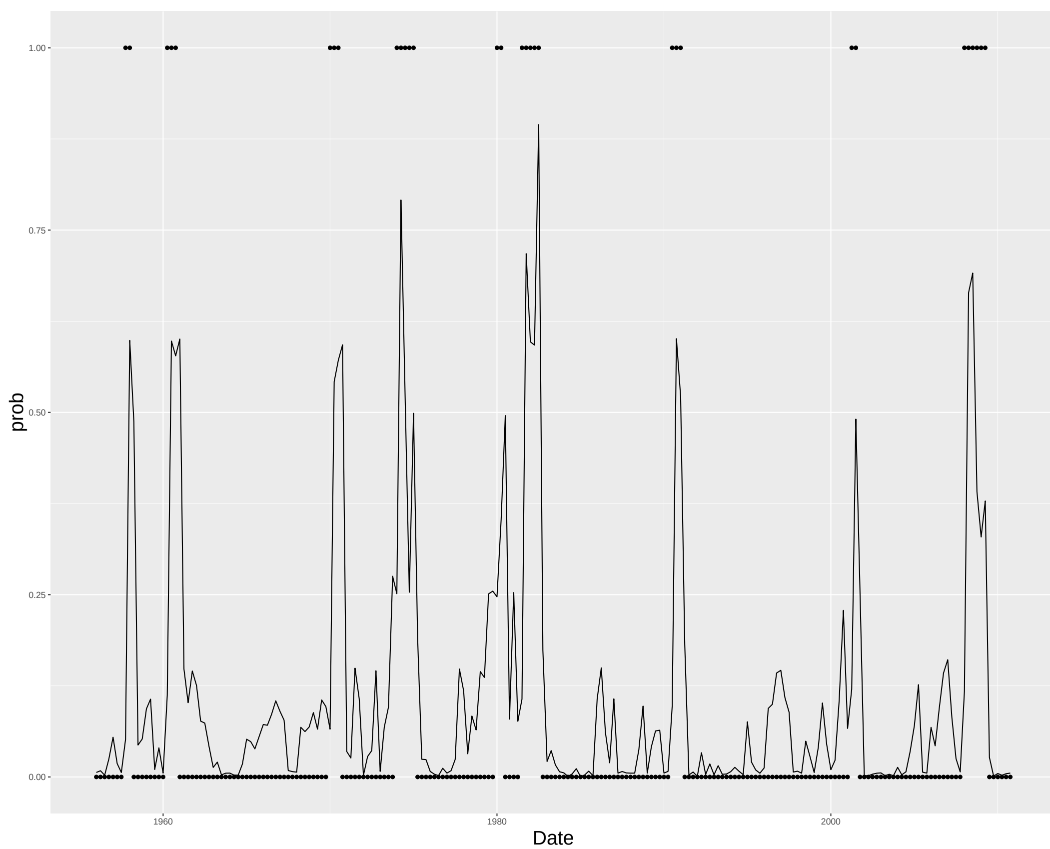

Figure 4.3: The line in the plot denotes the estimated $\Phi(\hat{\theta}(X^i, Z^i))$ by time. The black dots are the values of Y^i.

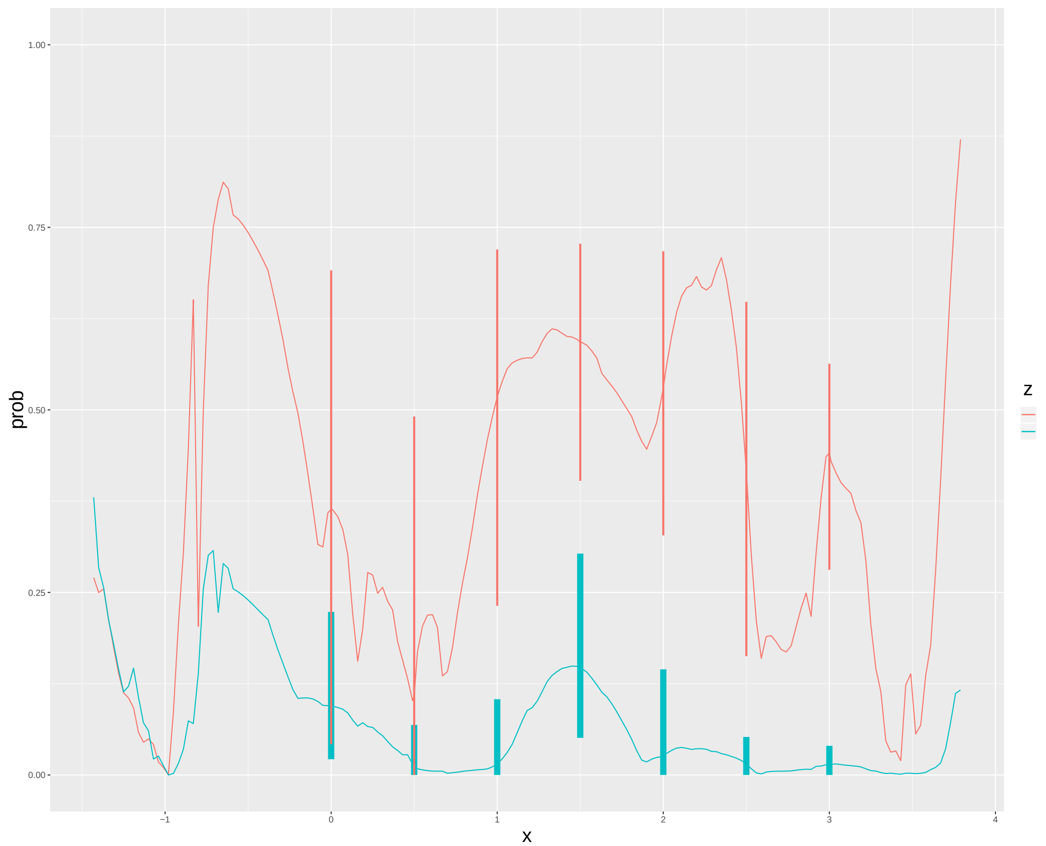

Figure 4.4: $\Phi(\hat{\theta}(x, z))$ and 95% confidence intervals based on the autoregression bootstrap.

Appendix

We give technical details in the appendix. We first summarize our notations.

- $(\mathbf{X}, \mathbf{Z}, Y)$ is the triple such that $(\mathbf{X}, \mathbf{Z}, Y) \overset{\mathrm{d}}{=} (\mathbf{X}^i, \mathbf{Z}^i, Y^i)$.

- α-mixing coefficient for stationary process $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : -\infty < i < \infty\}$ is defined by

$$\alpha(k) = \sup_{\substack{A \in \mathcal{F}_{-\infty}^{0} \\ B \in \mathcal{F}_{k}^{\infty}}} \left| P(A \cap B) - P(A)P(B) \right|, \ k \geq 1.$$

where $\mathcal{F}_a^b$ denotes the σ-field genereted by $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i) : a \leq i \leq b\}$.

- For $\boldsymbol{\beta} \in \mathbb{R}^\nu$,

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\beta}, y) = \frac{\partial \log f(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \in \mathbb{R}^\nu$$
$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\beta}, y) = \frac{\partial^2 \log f(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \in \mathbb{R}^{\nu \times \nu}$$

- $\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) = -E\left[\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{Z}), Y) | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}\right]$.

- $\mathbf{W}^i = (1, (X_1^i - x_1)/h_1, \ldots, (X_d^i - x_d)/h_d)^\top \in \mathbb{R}^{d+1}$, $1 \leq i \leq n$.

- $\mathbf{H} = \mathrm{diag}(h_1, \ldots, h_d) \in \mathbb{R}^{d \times d}$.

- $h_{\max} = \max_{1 \le j \le d} h_j$ and $\lambda_{\max} = \max_j \lambda_j$.

- $\tau_n = (nh_1 \times \cdots \times h_d / \log n)^{1/2}$ and $\upsilon_n = \tau_n(h_1 \times \cdots \times h_d \times \min_{1 \le j \le d} h_j)^{-1}$.

- $\mathrm{supp}(\mathbf{U})$ is the support of a random vector $\mathbf{U}$.

- $\mathbf{0}$ is the zero vector in a suitable vector space which may be different at each line.

- $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ is the identity matrix for $k \in \mathbb{N}$.

- $(L)_{\mathbf{g}}(\mathbf{v}, \mathbf{w}) = \frac{1}{g_1 \cdots g_d} L\left(\frac{v_1 - w_1}{g_1}\right) \cdots L\left(\frac{v_d - w_d}{g_d}\right)$ for $\mathbf{g} = (g_j)$, $\mathbf{v} = (v_j)$, $\mathbf{w} = (w_j) \in \mathbb{R}^d$ and a univariate function L.

- Define

$$\mathbf{M}(\mathbf{x}) = \int \mathbf{u}_n \mathbf{u}_n^\top \left(K^2\right)_{\mathbf{h}} (\mathbf{x}, \mathbf{u}) \mathrm{d}\mathbf{u} \in \mathbb{R}^{(d+1) \times (d+1)},$$
$$\mathbf{N}(\mathbf{x}) = \int \mathbf{u}_n \mathbf{u}_n^\top (K)_{\mathbf{h}} (\mathbf{x}, \mathbf{u}) \mathrm{d}\mathbf{u} \in \mathbb{R}^{(d+1) \times (d+1)}.$$

  where $\mathbf{u}_n = \left(1, \frac{u_1 - x_1}{h_1}, \ldots, \frac{u_d - x_d}{h_d}\right)^\top$.

- $\otimes$ denotes the Kronecker product. For two matrix $\mathbf{A} = (a_{jk})$ and $\mathbf{B}$,

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,m_2}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m_1,1}\mathbf{B} & \cdots & a_{m_1,m_2}\mathbf{B} \end{pmatrix}$$

- $\iota(\boldsymbol{\beta}, \mathbf{B}) = (\beta_1, \mathbf{B}_1, \ldots, \beta_\nu, \mathbf{B}_\nu)^\top \in \mathbb{R}^{\nu(d+1)}$ for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\nu)^\top \in \mathbb{R}^\nu$ and $\mathbf{B} \in \mathbb{R}^{\nu \times d}$ whose j-th row is $\mathbf{B}_j \in \mathbb{R}^{1 \times d}$, $j = 1, \ldots, \nu$.

- For $1 \leq p \leq \infty$, $\|\mathbf{u}\|_p$ denotes the ℓ_p-norm for a vector $\mathbf{u}$ and $\|\mathbf{U}\|_p$ denotes the matrix norm induced by the vector ℓ_p-norm for a matrix $\mathbf{U}$.

- $\|\mathbf{U}\|_{\mathrm{F}}$ denotes the Frobenius norm for a matrix $\mathbf{U}$.

- w.p.t.o. stands for with probability tending to one.

- (const.) denotes a generic positive constant which may attain different values at each line.

A.1 Proof of Theorem 2.1

Without loss of generality, we prove this theorem for $\nu = 2$ only, i.e. $\boldsymbol{\theta}(\mathbf{x}, \mathbf{z}) = (\theta_1, \theta_2)^\top(\mathbf{x}, \mathbf{z})$. Define a function space

$$\begin{aligned}\mathfrak{F} = \{\boldsymbol{\beta} = (\beta_{10}, \ldots, \beta_{1d}, \beta_{20}, \ldots, \beta_{2d})^\top : \beta_{kj}(\mathbf{x}, \mathbf{z}) \text{ is continuous in } \mathbf{x} \\ \text{for each } k, j \text{ and } \mathbf{z}\}\end{aligned}$$

and a norm on $\mathfrak{F}$ by

$$\|\boldsymbol{\beta}\| = \max_{k,j,\mathbf{z}} \sup_{\mathbf{x} \in [0,1]^d} |\beta_{kj}(\mathbf{x}, \mathbf{z})|$$

Note that $\mathfrak{F}$ with the above norm is a Banach space. Define a nonlinear map $\hat{\mathbf{A}} : \mathfrak{F} \to \mathfrak{F}$ by

$$\hat{\mathbf{A}}(\boldsymbol{\beta})(\mathbf{x}, \mathbf{z}) = n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \dot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\boldsymbol{\beta}(\mathbf{x}, \mathbf{z})), Y^i\right) \otimes \mathbf{W}^i$$

where

$$\mathbf{L}^i(\boldsymbol{\beta}(\mathbf{x},\mathbf{z})) = \left(\beta_{10}(\mathbf{x},\mathbf{z}) + \sum_{j=1}^{d} \beta_{1j}(\mathbf{x},\mathbf{z})\frac{X_j^i - x_j}{h_j}, \beta_{20}(\mathbf{x},\mathbf{z}) + \sum_{j=1}^{d} \beta_{2j}(\mathbf{x},\mathbf{z})\frac{X_j^i - x_j}{h_j}\right)^\top$$

Note that for each $(\mathbf{x},\mathbf{z})$, $\hat{\mathbf{A}}(\boldsymbol{\beta})(\mathbf{x},\mathbf{z}) = n^{-1}\partial L_n(\boldsymbol{\gamma},\boldsymbol{\Gamma};\mathbf{x},\mathbf{z})/\partial(\iota(\boldsymbol{\gamma},\boldsymbol{\Gamma}))$ where $\boldsymbol{\gamma} = (\beta_{10},\beta_{20})^\top(\mathbf{x},\mathbf{z})$ and $\boldsymbol{\Gamma} = (\beta_{kj}(\mathbf{x},\mathbf{z})) \in \mathbb{R}^{\nu\times d}$. $\hat{\mathbf{A}}$ has Gâteaux derivative

$$D\hat{\mathbf{A}}(\boldsymbol{\beta})(\boldsymbol{\delta})(\mathbf{x},\mathbf{z}) = n^{-1}\sum_{i=1}^{n} w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\boldsymbol{\beta}(\mathbf{x},\mathbf{z})), Y^i\right)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\boldsymbol{\delta}(\mathbf{x},\mathbf{z})$$

Moreover, $D\hat{\mathbf{A}}(\boldsymbol{\beta})$ is continuous in $\boldsymbol{\beta}$, which implies that $\hat{\mathbf{A}}$ is Fréchet differentiable and its Fréchet derivative is $D\hat{\mathbf{A}}(\boldsymbol{\beta})$.

Let $\tilde{\boldsymbol{\theta}} = (\theta_1, h_1\partial\theta_1/\partial x_1, \ldots, h_d\partial\theta_1/\partial x_d, \theta_2, h_1\partial\theta_2/\partial x_1, \ldots, h_d\partial\theta_2/\partial x_d)^\top \in \mathfrak{F}$ and $B_r(\tilde{\boldsymbol{\theta}})$ denote the r-neighborhood of $\tilde{\boldsymbol{\theta}}$. Note that $\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})) = \boldsymbol{\theta}(\mathbf{x},\mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x},\mathbf{z})(\mathbf{X}^i - \mathbf{x})$.

We apply Newton-Kantorovich theorem(Deimling (1985, Theorem 15.6)).

Proposition A.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces, F be a mapping $B_r(\zeta_0) \subset \mathcal{X} \to \mathcal{Y}$, where $B_r(\zeta_0)$ denotes a ball centered at ζ_0 with radius r, and $DF(\zeta)$ be the Fréchet derivative of F at ζ. Suppose that there exist constants α, β, γ and r such that $2\alpha\beta\gamma < 1$ and $2\alpha < r$ for which F has a derivative $DF(\zeta)$ for $\zeta \in B_r(\zeta_0)$, $DF(\zeta_0)$ is invertible, $\|DF(\zeta_0)^{-1}F(\zeta_0)\| \le \alpha$, $\|DF(\zeta_0)^{-1}\| \le \beta$, $\|DF(\zeta) - DF(\zeta')\| \le \gamma\|\zeta - \zeta'\|$ for all $\zeta, \zeta' \in B_r(\zeta_0)$.*
Then $F(\zeta) = 0$ has a unique solution ζ^* in $B_{2\alpha}(\zeta_0)$.

To apply Proposition A.1, we will show that w.p.t.o.

(a) $D\hat{\mathbf{A}}(\tilde{\boldsymbol{\theta}})$ is invertible and $\left\|\left(D\hat{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\right)^{-1}\right\| \leq (\text{const.})$

(b) $^{\exists}r > 0$ such that $\left\|D\hat{\mathbf{A}}(\boldsymbol{\beta}) - D\hat{\mathbf{A}}(\boldsymbol{\gamma})\right\| \leq (\text{const.}) \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|$ for all $\boldsymbol{\beta}, \boldsymbol{\gamma} \in B_r(\tilde{\boldsymbol{\theta}})$

(c) $\left\|\hat{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\right\| = O\left(\tau_n^{-1} + h_{\max}^2 + \lambda_{\max}\right).$

To prove (a), we first show that w.p.t.o.

$$\sup_{\mathbf{x},\mathbf{z}} \left\| n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^i\right) \otimes \left(\mathbf{W}^i (\mathbf{W}^i)^\top\right) \right.$$
$$\left. - E\left[w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^i\right) \otimes \left(\mathbf{W}^i (\mathbf{W}^i)^\top\right)\right] \right\|_{\mathrm{F}} = O(\tau_n^{-1}) \tag{A.2}$$

We only prove

$$\sup_{\mathbf{x},\mathbf{z}} \left| n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\ell}_{11}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^i\right) - E\left[w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\ell}_{11}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^i\right)\right] \right|$$
$$\overset{write}{=} \sup_{\mathbf{x},\mathbf{z}} |S_n(\mathbf{x},\mathbf{z})|$$
$$= O(\tau_n^{-1}) \text{ w.p.t.o.} \tag{A.3}$$

where $\ddot{\ell}_{11}$ is the $(1,1)$ entry of $\ddot{\boldsymbol{\ell}}$. The proof for the other elements in (A.2) is similar to that of (A.3).

To prove (A.3), we follow the lines of Masry (1996). Let $\mathbf{z} \in \mathrm{supp}(\mathbf{Z})$ be fixed. Decompose $[0,1]^d$ into $\{\mathcal{C}_l : \mathcal{C}_l$ is hypercube with center $\mathbf{x}_l$ and sides of $L_n^{-1/d},\ l = 1, \ldots, L_n\}$ where $L_n = O\left(v_n^d\right)$. Then,

$$\sup_{\mathbf{x}\in[0,1]^d} |S_n(\mathbf{x},\mathbf{z})| \leq \max_{1\leq l\leq L_n} \sup_{\mathbf{x}\in\mathcal{C}_l} |S_n(\mathbf{x},\mathbf{z}) - S_n(\mathbf{x}_l,\mathbf{z})| + \max_{1\leq l\leq L_n} |S_n(\mathbf{x}_l,\mathbf{z})| \tag{A.4}$$

We observe that

$$\max_{1\le l\le L_n}\sup_{\mathbf{x}\in\mathcal{C}_l}|S_n(\mathbf{x},\mathbf{z})-S_n(\mathbf{x}_l,\mathbf{z})| \le (\text{const.})(h_1\cdots h_d)^{-1}\max_{1\le j\le d}\left(h_j^{-1}\right)L_n^{-1/d}$$
$$=(\text{const.})\tau_n^{-1} \tag{A.5}$$

a.s.

To show $\max_{1\le l\le L_n}|S_n(\mathbf{x}_l,\mathbf{z})| = O\left(\tau_n^{-1}\right)$ w.p.t.o., decompose the sum $S_n(\mathbf{x}_l,\mathbf{z})$ into $2q_n = \lfloor n/\tau_n\rfloor$ blocks $V_1(\mathbf{x}_l),\ldots,V_{2q_n}(\mathbf{x}_l)$ with the same size τ_n and the remainder block $V_{2q_n+1}(\mathbf{x}_l)$ for each $1\le l\le L_n$. We will show that

$$\max_{1\le l\le L_n}\left|\sum_{j=1}^{q_n}V_{2j-1}(\mathbf{x}_l)\right| = O(\tau_n^{-1}) \text{ w.p.t.o.} \tag{A.6}$$

By Bradley (1983, Theorem 3), for each $\mathbf{x}$ and n, there exist independent random variables $\tilde{V}_{2j-1}(\mathbf{x})$, $j=1,\ldots,q_n$ such that for all $j=1,\ldots,q_n$, $\tilde{V}_{2j-1}(\mathbf{x})$ has the same distribution with $V_{2j-1}(\mathbf{x})$ and satisfies

$$P\left(\left|\tilde{V}_{2j-1}(\mathbf{x})-V_{2j-1}(\mathbf{x})\right|\ge\delta\right)\le 18\left(\|V_{2j-1}(\mathbf{x})\|_\infty/\delta\right)^{1/2}\alpha(\tau_n)$$

for all $\delta\le\|V_{2j-1}(\mathbf{x})\|_\infty$ where $\|\cdot\|_\infty$ denotes the essential supremum with respect to P.

For $C_2>0$, we have

$$\begin{aligned}
&P\left(\max_{1\le l\le L_n}\left|\sum_{j=1}^{q_n}V_{2j-1}(\mathbf{x}_l)\right|>C_2\tau_n^{-1}\right)\\
&\le L_n\max_{1\le l\le L_n}P\left(\left|\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)\right|>\frac{C_2\tau_n^{-1}}{2}\right)\\
&\quad+L_n\max_{1\le l\le L_n}P\left(\left|\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)-\sum_{j=1}^{q_n}V_{2j-1}(\mathbf{x}_l)\right|>\frac{C_2\tau_n^{-1}}{2}\right)
\end{aligned} \tag{A.7}$$

Since $\max_{1\le l\le L_n}\|V_{2j-1}(\mathbf{x}_l)\|_\infty = O\left(\tau_n(nh_1\cdots h_d)^{-1}\right)$, we obtain

$$\max_{1\le l\le L_n} P\left(\left|\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)-\sum_{j=1}^{q_n}V_{2j-1}(\mathbf{x}_l)\right| > \frac{C_2\tau_n^{-1}}{2}\right)$$
$$\le (\text{const.})q_n\left(\frac{\tau_n}{nh_1\cdots h_d}\frac{q_n\tau_n}{C_2}\right)^{1/2}\alpha(\tau_n) \tag{A.8}$$

Markov inequality and fundamental inequalities such as $e^x \le 1+x+x^2$ for $|x|\le 1/2$ lead to

$$\begin{aligned}
&P\left(\left|\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)\right| > C_2\tau_n^{-1}/2\right)\\
&\le P\left(\exp\left(\lambda_n\left|\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)\right|\right) > \exp(\lambda_n C_2\tau_n^{-1}/2)\right)\\
&\le e^{-\lambda_n C_2\tau_n^{-1}/2}\left[E\exp\left(\lambda_n\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)\right)\right.\\
&\quad\left.+E\exp\left(-\lambda_n\sum_{j=1}^{q_n}\tilde{V}_{2j-1}(\mathbf{x}_l)\right)\right]\\
&\le 2e^{-\lambda_n C_2\tau_n^{-1}/2}\exp\left(\lambda_n^2\sum_{j=1}^{q_n}E\tilde{V}_{2j-1}(\mathbf{x}_l)^2\right)
\end{aligned} \tag{A.9}$$

for $\lambda_n > 0$ satisfying $\lambda_n\max_{1\le l\le L_n,1\le j\le q_n}|\tilde{V}_{2j-1}(\mathbf{x}_l)| \le 1/2$. We can choose $\lambda_n = O\left(nh_1\cdots h_d/\tau_n\right)$. Write $S_n(\mathbf{x}) = \sum_{i=1}^n T_i(\mathbf{x})$. By stationarity of $(\mathbf{X}^i,\mathbf{Z}^i,Y^i)$, we have

$$\begin{aligned}
\sum_{j=1}^{q_n}EV_{2j-1}(\mathbf{x}_l)^2 &\le (\text{const.})\left[n^{-2}\sum_{i=1}^n E\left(w_{\mathrm{c}}^i(\mathbf{x}_l)\right)^2+\sum_{i\ne i'}\left|\mathrm{Cov}\left(T_i(\mathbf{x}_l),T_{i'}(\mathbf{x}_l)\right)\right|\right]\\
&\le (\text{const.})\left[(nh_1\cdots h_d)^{-1}+n\sum_{i=2}^n\left|\mathrm{Cov}\left(T_1(\mathbf{x}_l),T_i(\mathbf{x}_l)\right)\right|\right]
\end{aligned} \tag{A.10}$$

Let $\pi_n = (h_1 \cdots h_d)^{b^{-1}(2/a-1)}$. By Hall et al. (2014, Corollary A.2) and (C9), we obtain

$$\begin{aligned}
& n \sum_{i=2}^{n} |\mathrm{Cov}\,(T_1(\mathbf{x}_l), T_i(\mathbf{x}_l))| \\
&= n \sum_{i=2}^{\pi_n} |\mathrm{Cov}\,(T_1(\mathbf{x}_l), T_i(\mathbf{x}_l))| + n \sum_{i=\pi_n+1}^{n} |\mathrm{Cov}\,(T_1(\mathbf{x}_l), T_i(\mathbf{x}_l))| \\
&\leq (\mathrm{const.}) \left[\frac{\pi_n}{n} + n \sum_{i=\pi_n+1}^{n} \|T_1(\mathbf{x}_l)\|_a^2 \, \alpha(i)^{1-2/a} \right] \\
&\leq (\mathrm{const.}) \left[\frac{\pi_n}{n} + n^{-1} \sum_{i=\pi_n+1}^{n} \left(\frac{i}{\pi_n} \right)^b (h_1 \cdots h_d)^{-2(a-1)/a} \alpha(i)^{1-2/a} \right] \\
&\leq (\mathrm{const.}) \left[\frac{\pi_n}{n} + n^{-1} \pi_n^{-b} (h_1 \cdots h_d)^{-2(a-1)/a} \right] \\
&= O\left((n h_1 \cdots h_d)^{-1} \right) \qquad \text{(A.11)}
\end{aligned}$$

where $\|T_1(\mathbf{x}_l)\|_a^a$ is a-th moment of $T_1(\mathbf{x}_l)$. (A.7), (A.8), (A.9), (A.10) and (A.11) imply

$$\begin{aligned}
& P\left(\max_{1 \leq l \leq L_n} \left| \sum_{j=1}^{q_n} V_{2j-1}(\mathbf{x}_l) \right| > C_2 \tau_n^{-1} \right) \\
&\leq L_n \exp\left(-\lambda_n C_2 \tau_n^{-1}/2 + (\mathrm{const.}) \lambda_n^2 (n h_1 \times \cdots \times h_d)^{-1} \right) \\
&\quad + (\mathrm{const.}) L_n q_n \left(\frac{\tau_n}{n h_1 \times \cdots \times h_d} \frac{q_n \tau_n}{C_2} \right)^{1/2} \alpha(\tau_n) \\
&= L_n n^{-(\mathrm{const.})C_2 + (\mathrm{const.})} + (\mathrm{const.}) C_2^{-1/2} \cdot n \upsilon_n^d \left(\tau_n h_1 \times \cdots \times h_d \right)^{1/2} \alpha(\tau_n)
\end{aligned}$$

By (C7), $L_n n^{-(\mathrm{const.})C_2 + (\mathrm{const.})} \to 0$ as $n \to \infty$ for sufficiently large $C_2 > 0$. Since we have $\upsilon_n \to \infty$ as $n \to \infty$ by (C6) and (C7), we also have $n \upsilon_n^d \left(\tau_n h_1 \times \cdots \times h_d \right)^{1/2} \alpha(\tau_n) \to 0$ from (C8). These imply (A.6). In the same manner, we can show that $\max_{1\le l\le L_n}\left|\sum_{j=1}^{q_n} V_{2j}(\mathbf{x}_l)\right|$ and $\max_{1\le l\le L_n}|V_{2q_n+1}(\mathbf{x}_l)|$ have the same order of magnitudes. Since $\mathbf{z}$ was arbitrary and $\mathrm{supp}(\mathbf{Z})$ is a finite set, we have (A.3).

Define a bounded linear operator on $\mathfrak{F}$ by

$$D\mathbf{A}_{\boldsymbol{\theta}}(\boldsymbol{\delta})(\mathbf{x},\mathbf{z}) = -p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})\left[\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})\otimes\mathbf{N}(\mathbf{x})\right]\boldsymbol{\delta}(\mathbf{x},\mathbf{z}).$$

We now approximate $E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})),Y^i\right)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\right]$. We get uniformly for $\mathbf{x},\mathbf{z}$,

$$\begin{aligned}
&E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})),Y^i\right)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\right]\\
&=E\left[E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})),Y^i\right)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)|\mathbf{X}^i,\mathbf{Z}^i\right]\right]\\
&=E\Bigg[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\sum_y\left\{\ddot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{x},\mathbf{z})+\boldsymbol{\Theta}(\mathbf{x},\mathbf{z})(\mathbf{X}^i-\mathbf{x}),y\right)f(y,\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i))\right\}\\
&\quad\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\Bigg]\\
&=E\Bigg[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\sum_y\left\{\ddot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i),y\right)f(y,\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i))\right\}\\
&\quad\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\Bigg]+O\left(h_{\max}^2+\lambda_{\max}\right)\\
&=-E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\otimes\left(\mathbf{W}^i(\mathbf{W}^i)^\top\right)\right]+O\left(h_{\max}^2+\lambda_{\max}\right)
\end{aligned}$$

In the third equality above, we have used $\max_{\mathbf{z},1\le i\le n} w_{\mathrm{d}}^i(\mathbf{z})\le 1$, (C12) and the fact that $\mathrm{supp}(Y)$ is finite. Furthermore, combining the fact that $\mathrm{supp}(\mathbf{Z})$ is finite, (C3), (C6) and (C13), we obtain following approximation

$$\sup_{\mathbf{x},\mathbf{z}} \Bigg\| - p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}) \left[\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}) \otimes \mathbf{N}(\mathbf{x})\right] \\ - E\left[w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \ddot{\boldsymbol{\ell}}\left(\mathbf{L}^i(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^i\right) \otimes \left(\mathbf{W}^i (\mathbf{W}^i)^\top\right)\right] \Bigg\|_{\mathrm{F}} = o(1)$$

This and (A.2) imply

$$\left\| D\hat{\mathbf{A}}(\tilde{\boldsymbol{\theta}}) - D\mathbf{A}_{\boldsymbol{\theta}} \right\| = o_P(1) \tag{A.12}$$

To complete the proof of (a), we use the following facts (F1), (F2) and (F3).

(F1) If both $\mathbf{B}_1 \in \mathbb{R}^{m_1 \times m_1}$ and $\mathbf{B}_2 \in \mathbb{R}^{m_2 \times m_2}$ are invertible, then $(\mathbf{B}_1 \otimes \mathbf{B}_2)^{-1} = \mathbf{B}_1^{-1} \otimes \mathbf{B}_2^{-1}$.

(F2) If $\mathbf{B}_1 \in \mathbb{R}^{m_1 \times m_1}$ has eigen values $\{\mu_{1,j} : 1 \leq j \leq m_1\}$ and $\mathbf{B}_2 \in \mathbb{R}^{m_2 \times m_2}$ has eigen values $\{\mu_{2,j} : 1 \leq j \leq m_2\}$, then $\mathbf{B}_1 \otimes \mathbf{B}_2$ has eigen values $\{\mu_{1,j} \times \mu_{2,j'} : 1 \leq j \leq m_1, 1 \leq j' \leq m_2\}$.

(F3) Let A_1 and A_2 be bounded linear operators on a Banach space with a norm $\|\cdot\|_{\mathrm{A}}$. If A_1 is invertible and $\|A_1 - A_2\|_{\mathrm{A}} < \varepsilon / \left\|A_1^{-1}\right\|_{\mathrm{A}}$ for some $0 < \varepsilon < 1$, then A_2 is invertible and $A_2^{-1} = A_1^{-1} \sum_{n=0}^{\infty} (I - A_2 A_1^{-1})^n$ with $\left\|A_2^{-1}\right\|_{\mathrm{A}} \leq \left\|A_1^{-1}\right\|_{\mathrm{A}} / (1 - \varepsilon)$.

Utilizing (F1) and (F2), we can show that $D\mathbf{A}_{\boldsymbol{\theta}}$ is a invertible operator for all n and $\sup_n \left\| D\mathbf{A}_{\boldsymbol{\theta}}^{-1} \right\| < \infty$, which with (F3) and (A.12) completes the proof of (a).

By (C11) and (C12), we can choose $r > 0$ such that for $\boldsymbol{\beta}, \boldsymbol{\gamma} \in B_r(\tilde{\boldsymbol{\theta}})$ and $\|\boldsymbol{\delta}\| \le 1$

$$\begin{aligned}
&\left\| D\hat{\mathbf{A}}(\boldsymbol{\beta})(\boldsymbol{\delta}) - D\hat{\mathbf{A}}(\boldsymbol{\gamma})(\boldsymbol{\delta}) \right\| \\
&\le (\text{const.}) \|\boldsymbol{\delta}\| \cdot \sup_{\mathbf{x},\mathbf{z}} \Bigg\| n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \\
&\quad \times \left[\ddot{\boldsymbol{\ell}}\left(\mathbf{L}^{i}(\boldsymbol{\beta}(\mathbf{x},\mathbf{z})), Y^{i}\right) - \ddot{\boldsymbol{\ell}}\left(\mathbf{L}^{i}(\boldsymbol{\gamma}(\mathbf{x},\mathbf{z})), Y^{i}\right) \right] \otimes \mathbf{W}^{i}(\mathbf{W}^{i})^{\top} \Bigg\|_{\max} \\
&\le (\text{const.}) \|\boldsymbol{\beta} - \boldsymbol{\gamma}\| \cdot \sup_{\mathbf{x},\mathbf{z}} n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z})
\end{aligned}$$

a.s. where $\|\mathbf{B}\|_{\max} = \max_{ij} |b_{ij}|$ for a matrix $\mathbf{B} = (b_{ij})$. Moreover, we can prove $\sup_{\mathbf{x},\mathbf{z}} n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) = O(1)$ w.p.t.o. in a similar way to the proof of (A.3). Consequently, we get (b).

To prove (c), write

$$\begin{aligned}
&n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \dot{\boldsymbol{\ell}}\left(\mathbf{L}^{i}(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^{i}\right) \otimes \mathbf{W}^{i} \\
&= n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \left[\dot{\boldsymbol{\ell}}\left(\mathbf{L}^{i}(\tilde{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})), Y^{i}\right) - \dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^{i}, \mathbf{Z}^{i}), Y^{i}\right) \right] \otimes \mathbf{W}^{i} \\
&\quad + n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^{i}, \mathbf{Z}^{i}), Y^{i}\right) \otimes \mathbf{W}^{i} \\
&= \mathbf{S}_{n,1}(\mathbf{x},\mathbf{z}) + \mathbf{S}_{n,2}(\mathbf{x},\mathbf{z})
\end{aligned}$$

In a similar way to the proof of (A.2), we can show that $\|\mathbf{S}_{nj} - E\mathbf{S}_{nj}\| = O(\tau_n^{-1})$ w.p.t.o. for $j = 1, 2$. Moreover, we have $\|E\mathbf{S}_{n,1}\| = O(h_{\max}^2 + \lambda_{\max})$ and $E\mathbf{S}_{n,2} = \mathbf{0}$. Therefore, we obtain (c).

By (a), (b) and (c), we can apply Proposition A.1 and obtain the desired results: $^\exists\hat{\boldsymbol{\theta}} \in \mathfrak{F}$ such that $\hat{\mathbf{A}}\left(\hat{\boldsymbol{\theta}}\right) = \mathbf{0}$ and $\left\|\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right\| = O\left(\tau_n^{-1} + h_{\max}^2 + \lambda_{\max}\right)$ w.p.t.o.

A.2 Proof of Lemma 2.2

Let $(\mathbf{x}, \mathbf{z}) \in \mathrm{supp}(\mathbf{X}, \mathbf{Z})$ be fixed. Define

$$\begin{aligned}
\hat{\mathbf{C}}(\boldsymbol{\beta}, \mathbf{B}) &= n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \\
&\quad \times \dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{x}, \mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})(\mathbf{X}^i - \mathbf{x}) + \boldsymbol{\beta} + \mathbf{B}\mathbf{H}^{-1}(\mathbf{X}^i - \mathbf{x}), Y^i\right) \otimes \mathbf{W}^i \\
\hat{\mathbf{D}}(\boldsymbol{\beta}, \mathbf{B}) &= n^{-1} \sum_{i=1}^{n} w_{\mathrm{c}}^{i}(\mathbf{x}) w_{\mathrm{d}}^{i}(\mathbf{z}) \\
&\quad \times \ddot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{x}, \mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})(\mathbf{X}^i - \mathbf{x}) + \boldsymbol{\beta} + \mathbf{B}\mathbf{H}^{-1}(\mathbf{X}^i - \mathbf{x}), Y^i\right) \otimes \left(\mathbf{W}^i (\mathbf{W}^i)^\top\right).
\end{aligned}$$

for $\boldsymbol{\beta} \in \mathbb{R}^\nu$ and $\mathbf{B} \in \mathbb{R}^{\nu \times d}$. Note that $\hat{\mathbf{D}}(\boldsymbol{\beta}, \mathbf{B}) = \partial \hat{\mathbf{C}}(\boldsymbol{\beta}, \mathbf{B})^\top / \partial(\iota(\boldsymbol{\beta}, \mathbf{B}))$.
Let $\hat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}) = \hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) - \boldsymbol{\theta}(\mathbf{x}, \mathbf{z})$ and $\hat{\mathbf{B}}(\mathbf{x}, \mathbf{z}) = \left[\hat{\boldsymbol{\Theta}}(\mathbf{x}, \mathbf{z}) - \boldsymbol{\Theta}(\mathbf{x}, \mathbf{z})\right] \mathbf{H}$. We have $\hat{\mathbf{C}}\left(\hat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}), \hat{\mathbf{B}}(\mathbf{x}, \mathbf{z})\right) = \mathbf{0}$ by Theorem 2.1. There exists $\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)$ such that $\iota\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)$ lies in between $\iota\left(\mathbf{0}, \mathbf{O}\right)$ and $\iota\left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{B}}\right)$ and satisfies

$$\mathbf{0} = \hat{\mathbf{C}}\left(\hat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}), \hat{\mathbf{B}}(\mathbf{x}, \mathbf{z})\right) = \hat{\mathbf{C}}\left(\mathbf{0}, \mathbf{O}\right) + \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right) \iota\left(\hat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{z}), \hat{\mathbf{B}}(\mathbf{x}, \mathbf{z}\right) \quad \text{(A.13)}$$

where $\mathbf{O} \in \mathbb{R}^{\nu \times d}$ is the zero matrix.

We now discuss the properties of $\hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)$. We first approximate $E\hat{\mathbf{D}}\left(\mathbf{0}, \mathbf{O}\right)$. As in the proof of Theorem 2.1, we can show

$$\left\| E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O}) + p_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) \mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) \otimes \mathbf{N}(\mathbf{x}) \right\|_2 = o(1) \quad \text{(A.14)}$$

(A.14), (F3) (in the proof of Theorem 2.1), conditions (C3) and (C13) yield that $E\hat{\mathbf{D}}(\mathbf{0},\mathbf{O})$ is invertible and there exists $C>0$ such that

$$\left\|E\hat{\mathbf{D}}(\mathbf{0},\mathbf{O})^{-1}\right\|_2 < (\text{const.})\left\|[p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})\otimes\mathbf{N}(\mathbf{x})]^{-1}\right\|_2 < C$$

for sufficiently large n.

Let $0<\varepsilon<(3C)^{-1}$. Since $\iota(\boldsymbol{\beta},\mathbf{B})\mapsto E\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})$ is continuous in some neighborhood of $\iota(\mathbf{0},\mathbf{O})$ for sufficiently large n, we can choose $\delta>0$ such that

$$\left\|E\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})-E\hat{\mathbf{D}}(\mathbf{0},\mathbf{O})\right\|_2<\varepsilon \text{ for all } \|\iota(\boldsymbol{\beta},\mathbf{B})\|_2\le\delta \tag{A.15}$$

for sufficiently large n.

In a similar way to the proof of (A.2), we can show that

$$\sup_{\|\iota(\boldsymbol{\beta},\mathbf{B})\|_2\le\delta}\left\|\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})-E\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})\right\|_2=o_P(1). \tag{A.16}$$

We claim that (A.14), (A.15) and (A.16) imply that

$$\hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*,\hat{\mathbf{B}}^*\right)^{-1}=E\hat{\mathbf{D}}(\mathbf{0},\mathbf{O})^{-1}+o_P(1) \tag{A.17}$$

Then, (A.13) and (A.14) lead to

$$\begin{aligned}\iota\left(\hat{\boldsymbol{\beta}}(\mathbf{x},\mathbf{z}),\hat{\mathbf{B}}(\mathbf{x},\mathbf{z}\right)&=\left[-E\hat{\mathbf{D}}(\mathbf{0},\mathbf{O})^{-1}+o_P(1)\right]\hat{\mathbf{C}}\left(\mathbf{0},\mathbf{O}\right)\\&=\left[\mathbf{G}(\mathbf{x},\mathbf{z})^{-1}+o_P(1)\right]\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})\end{aligned}$$

To prove (A.17), define a event

$$\Omega_n=\left\{\left\|\iota\left(\hat{\boldsymbol{\beta}}(\mathbf{x},\mathbf{z}),\hat{\mathbf{B}}(\mathbf{x},\mathbf{z})\right)\right\|_2\le\delta,\ \sup_{\|\iota(\boldsymbol{\beta},\mathbf{B})\|_2\le\delta}\left\|\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})-E\hat{\mathbf{D}}(\boldsymbol{\beta},\mathbf{B})\right\|_2\le\varepsilon\right\}$$

By Theorem 2.1 and (A.16), $P(\Omega_n) \to 1$ as $n \to 0$. Moreover, by (A.15), we have

$$\left\| \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right) - E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O}) \right\|_2 < 2\varepsilon < \frac{2}{3} \left\| E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O})^{-1} \right\|_2^{-1}$$

on Ω_n for sufficiently large n. This and (F3) yield that

- $\hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)$ is invertible,
- $\left\| \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)^{-1} \right\|_2 < (\text{const.}) \left\| E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O})^{-1} \right\|_2$.

on Ω_n for sufficiently large n. Then, we have (A.17) from

$$\begin{aligned} &\left\| \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)^{-1} - E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O})^{-1} \right\|_2 \\ &\leq \left\| \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right) - E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O}) \right\|_2 \times \left\| \hat{\mathbf{D}}\left(\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{B}}^*\right)^{-1} \right\|_2 \cdot \left\| E\hat{\mathbf{D}}(\mathbf{0}, \mathbf{O})^{-1} \right\|_2 \\ &< (\text{const.}) C^2 \varepsilon. \end{aligned}$$

A.3 Proof of Lemma 2.3

Let $(\mathbf{x}, \mathbf{z})$ be fixed. We write $\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z}) = n^{-1} \sum_{i=1}^n w_{\mathrm{c}}^i(\mathbf{x}) w_{\mathrm{d}}^i(\mathbf{z}) \mathbf{U}^i$. Then, as in the proof of (A.11), we can prove

$$\mathrm{Var}\left(\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z})\right) = n^{-2} \sum_{i=1}^n \mathrm{Var}\left(w_{\mathrm{c}}^i w_{\mathrm{d}}^i \mathbf{U}^i\right) + o\left((nh_1 \times \cdots \times h_d)^{-1}\right).$$

Moreover, standard calculations yield that

$$\begin{aligned} &n^{-2} \sum_{i=1}^n \mathrm{Var}\left(w_{\mathrm{c}}^i w_{\mathrm{d}}^i \mathbf{U}^i\right) \\ &= (nh_1 \times \cdots \times h_d)^{-1} \left[p_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) \mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) \otimes \mathbf{M}(\mathbf{x}) + o(1)\right]. \end{aligned}$$

Let $\tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z}) = \boldsymbol{\theta}(\mathbf{x},\mathbf{z}) + \boldsymbol{\Theta}(\mathbf{x},\mathbf{z})(\mathbf{X}^i - \mathbf{x})$. Define $\rho(\mathbf{v},\mathbf{w}) = \sum_j I(v_j \neq w_j)$ for $\mathbf{v} = (v_j)$ and $\mathbf{w} = (w_j)$. We get the approximation of the bias part of $\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})$. First, approximate and decompose

$$
\begin{aligned}
&E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\dot{\boldsymbol{\ell}}\left(\tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z}), Y^i\right)\right] \\
&= E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\left[\dot{\boldsymbol{\ell}}\left(\tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z}), Y^i\right) - \dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i), Y^i\right)\right]\right] \\
&= E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\ddot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i), Y^i\right)\left[\tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z}) - \boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i)\right]\right] + o(h_{\max}^2) \\
&= E\left[w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\left[\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i) - \tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z})\right]\right] + o(h_{\max}^2) \\
&= E\left[I(\mathbf{Z}^i = \mathbf{z})w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\left[\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i) - \tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z})\right]\right] \\
&\quad + E\left[I(\mathbf{Z}^i \neq \mathbf{z})w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\left[\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i) - \tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z})\right]\right] + o(h_{\max}^2) \\
&= \mathbf{S}_{01} + \mathbf{S}_{02} + o(h_{\max}^2)
\end{aligned}
$$

Let $\tilde{\mathbf{u}}_n = \left(\frac{u_1 - x_1}{h_1}, \ldots, \frac{u_d - x_d}{h_d}\right)^\top$. We obtain

$$
\begin{aligned}
\mathbf{S}_{01} =& \frac{1}{2}p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}) \\
&\times \left(\mathrm{tr}\left(\mathbf{H}\ddot{\boldsymbol{\theta}}_1(\mathbf{x},\mathbf{z})\mathbf{H}\int \tilde{\mathbf{u}}_n\tilde{\mathbf{u}}_n^\top\left(K\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\right),\right. \\
&\qquad \left.\ldots, \mathrm{tr}\left(\mathbf{H}\ddot{\boldsymbol{\theta}}_\nu(\mathbf{x},\mathbf{z})\mathbf{H}\int \tilde{\mathbf{u}}_n\tilde{\mathbf{u}}_n^\top\left(K\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\right)\right)^\top \\
&+ o(h_{\max}^2)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{S}_{02} =& E\left[I(\rho(\mathbf{Z}^i,\mathbf{z}) = 1)w_{\mathrm{c}}^i(\mathbf{x})w_{\mathrm{d}}^i(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\left[\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i) - \tilde{\boldsymbol{\theta}}^i(\mathbf{x},\mathbf{z})\right]\right] + o(\lambda_{\max}) \\
=& \int K_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\sum_j \lambda_j \sum_{\substack{\mathbf{z}':\rho(\mathbf{z}',\mathbf{z})=1,\\ z'_j \neq z_j}} p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}')\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}')\left[\boldsymbol{\theta}(\mathbf{x},\mathbf{z}') - \boldsymbol{\theta}(\mathbf{x},\mathbf{z})\right] + o(\lambda_{\max})
\end{aligned}
$$

In a similar way, we obatin

$$
\begin{aligned}
&E\left[w_{\mathrm{c}}^{i}(\mathbf{x})w_{\mathrm{d}}^{i}(\mathbf{z})\dot{\boldsymbol{\ell}}\left(\tilde{\boldsymbol{\theta}}^{i}(\mathbf{x},\mathbf{z}),Y^{i}\right)\frac{X_{j}^{i}-x_{j}}{h_{j}}\right]\\
&=E\left[I(\mathbf{Z}^{i}=\mathbf{z})w_{\mathrm{c}}^{i}(\mathbf{x})w_{\mathrm{d}}^{i}(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^{i},\mathbf{Z}^{i})\left[\boldsymbol{\theta}(\mathbf{X}^{i},\mathbf{Z}^{i})-\tilde{\boldsymbol{\theta}}^{i}(\mathbf{x},\mathbf{z})\right]\frac{X_{j}^{i}-x_{j}}{h_{j}}\right]\\
&\quad+E\left[I(\mathbf{Z}^{i}\neq\mathbf{z})w_{\mathrm{c}}^{i}(\mathbf{x})w_{\mathrm{d}}^{i}(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^{i},\mathbf{Z}^{i})\left[\boldsymbol{\theta}(\mathbf{X}^{i},\mathbf{Z}^{i})-\tilde{\boldsymbol{\theta}}^{i}(\mathbf{x},\mathbf{z})\right]\frac{X_{j}^{i}-x_{j}}{h_{j}}\right]+O(h_{\max}^{3})\\
&=\mathbf{S}_{j1}+\mathbf{S}_{j2}+O(h_{\max}^{3})
\end{aligned}
$$

for $j=1,\ldots,d$. For $\mathbf{S}_{j1}$, we approximate

$$
\begin{aligned}
\mathbf{S}_{j1}=&\frac{1}{2}p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z})\\
&\times\Bigg(\mathrm{tr}\left(\mathbf{H}\ddot{\boldsymbol{\theta}}_{1}(\mathbf{x},\mathbf{z})\mathbf{H}\int\frac{u_{j}-x_{j}}{h_{j}}\tilde{\mathbf{u}}_{n}\tilde{\mathbf{u}}_{n}^{\top}\left(K\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\right),\\
&\qquad\ldots,\mathrm{tr}\left(\mathbf{H}\ddot{\boldsymbol{\theta}}_{\nu}(\mathbf{x},\mathbf{z})\mathbf{H}\int\frac{u_{j}-x_{j}}{h_{j}}\tilde{\mathbf{u}}_{n}\tilde{\mathbf{u}}_{n}^{\top}\left(K\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\right)\Bigg)^{\top}\\
&+O(h_{\max}^{3}).
\end{aligned}
$$

Note that $\int\frac{u_{j}-x_{j}}{h_{j}}\tilde{\mathbf{u}}_{n}\tilde{\mathbf{u}}_{n}^{\top}\left(K\right)_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}$ is the zero-matrix if $\mathbf{x}$ is a interior point of $\mathrm{supp}(\mathbf{X})$. For $\mathbf{S}_{j2}$,

$$
\begin{aligned}
\mathbf{S}_{j2}=&E\left[I(\rho(\mathbf{Z}^{i},\mathbf{z})=1)w_{\mathrm{c}}^{i}(\mathbf{x})w_{\mathrm{d}}^{i}(\mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^{i},\mathbf{Z}^{i})\left[\boldsymbol{\theta}(\mathbf{X}^{i},\mathbf{Z}^{i})-\tilde{\boldsymbol{\theta}}^{i}(\mathbf{x},\mathbf{z})\right]\frac{X_{j}^{i}-x_{j}}{h_{j}}\right]+o(\lambda_{\max})\\
=&\sum_{j}\lambda_{j}\int\frac{u_{j}-x_{j}}{h_{j}}K_{\mathbf{h}}(\mathbf{x},\mathbf{u})d\mathbf{u}\sum_{\substack{\mathbf{z}':\rho(\mathbf{z}',\mathbf{z})=1,\\ z_{j}'\neq z_{j}}}p_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}')\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}')\left[\boldsymbol{\theta}(\mathbf{x},\mathbf{z}')-\boldsymbol{\theta}(\mathbf{x},\mathbf{z})\right]\\
&+o(\lambda_{\max}).
\end{aligned}
$$

By combining these approximations and applying Masry and Fan (1997), we get the asymptotic distribution of $\hat{\mathbf{F}}(\mathbf{x},\mathbf{z})$.

A.4 Proof of Theorem 3.1 and Theorem 3.2

Let $(\mathbf{x}, \mathbf{z})$ be fixed. Let

$$\boldsymbol{\Sigma}_\infty = p_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) \otimes \mathbf{M}(\mathbf{x})$$

As we have proven in Lemma 2.3, we have $\mathrm{Var}((nh_1 \times \cdots \times h_d)^{1/2}\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z})) = \boldsymbol{\Sigma}_\infty + o(1)$.

Since the distribution function of $N(\mathbf{0}, \boldsymbol{\Sigma}_\infty)$ is continuous, we obtain

$$d_{\mathrm{K}}\left(\mathcal{L}\left((nh_1 \times \cdots \times h_d)^{1/2}\hat{\mathbf{F}}(\mathbf{x}, \mathbf{z})\right), N(\mathbf{0}, \boldsymbol{\Sigma}_\infty)\right) = o(1)$$

by Polya's theorem(Bhattacharya and Rao (1986)), (U1) and Lemma 2.3. Moreover, we have $\hat{\mathbf{G}}(\mathbf{x}, \mathbf{z}) = \mathbf{G}(\mathbf{x}, \mathbf{z}) + o_P(1)$. Therefore, it is enough to show that

$$d_{\mathrm{K}}\left(\mathcal{L}^*\left((nh_1 \times \cdots \times h_d)^{1/2}\hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z})\right), N(\mathbf{0}, \boldsymbol{\Sigma}_\infty)\right) = o_P(1) \tag{A.18}$$

to prove Theorem 3.1 and Theorem 3.2. Write $(nh_1 \times \cdots \times h_d)^{1/2}\hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \boldsymbol{\chi}^{i,*}$. Let E^* and Var^* denote the conditional expectation and variance given $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$, $1 \le i \le n$, respectively.

We first prove Theorem 3.1. Note that $E^*\hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) = 0$ w.p.t.o. by (C12) and Theorem 2.1. Combining (C5), (C12), $\max_{1\le i\le n} w_{\mathrm{d}}^i(\mathbf{z}) \le 1$, and $|\mathrm{supp}(Y)| < \infty$, we have $\max_{1\le i\le n} \|\boldsymbol{\chi}^{i,*}\|_2 = O\left((nh_1 \times \cdots \times h_d)^{-1/2}\right)$ if $\sup_{\mathbf{x},\mathbf{z}} \left\|\hat{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) - \boldsymbol{\theta}(\mathbf{x}, \mathbf{z})\right\|_2$

is sufficiently close to 0. By Theorem 2.1, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{i=1}^{n} E^{*}\left[\left\|\boldsymbol{\chi}^{i,*}\right\|_2^2 I\left(\left\|\boldsymbol{\chi}^{i,*}\right\|_2 > \varepsilon\right)\right] \\
&= \sum_{i=1}^{n}\sum_{y}\left\|\boldsymbol{\chi}^{i,*}\right\|_2^2 I\left(\left\|\boldsymbol{\chi}^{i,*}\right\|_2 > \varepsilon\right) f(y, \hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^i)) = 0
\end{aligned}$$

w.p.t.o. Therefore, we obtain

$$^{\forall}\varepsilon > 0,\ \sum_{i=1}^{n} E^{*}\left[\left\|\boldsymbol{\chi}^{i,*}\right\|_2^2 I\left(\left\|\boldsymbol{\chi}^{i,*}\right\|_2 > \varepsilon\right)\right] = o_P(1) \tag{A.19}$$

From (C3), (C12), (C13) and Theorem 2.1, we have

$$\begin{aligned}
&\mathrm{Var}^{*}\left((nh_1 \times \cdots \times h_d)^{1/2}\hat{\mathbf{F}}^{*}(\mathbf{x},\mathbf{z})\right) \\
&= n^{-1}\sum_{i=1}^{n}\left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{X}^i) w_{\mathrm{d}}^i(\mathbf{z})^2 \\
&\quad \times E^{*}\left[\dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i), Y^{i,*}\right)\dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i), Y^{i,*}\right)^{\top}\right] \otimes \mathbf{W}^i(\mathbf{W}^i)^{\top} \\
&= n^{-1}\sum_{i=1}^{n}\left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{X}^i) w_{\mathrm{d}}^i(\mathbf{z})^2 \\
&\quad \times \left[\sum_{y}\dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i), y\right)\dot{\boldsymbol{\ell}}\left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i), y\right)^{\top} f\left(y, \hat{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i)\right)\right] \otimes \mathbf{W}^i(\mathbf{W}^i)^{\top} \\
&= n^{-1}\sum_{i=1}^{n}\left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{X}^i) w_{\mathrm{d}}^i(\mathbf{z})^2 \\
&\quad \times \left[\sum_{y}\dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i), y\right)\dot{\boldsymbol{\ell}}\left(\boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i), y\right)^{\top} f\left(y, \boldsymbol{\theta}(\mathbf{X}^i,\mathbf{Z}^i)\right)\right] \otimes \mathbf{W}^i(\mathbf{W}^i)^{\top} + O_P(\tau_n^{-1}) \\
&= n^{-1}\sum_{i=1}^{n}\left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{X}^i) w_{\mathrm{d}}^i(\mathbf{z})^2 \cdot \mathbf{I}_{\boldsymbol{\theta}}(\mathbf{X}^i,\mathbf{Z}^i) \otimes \mathbf{W}^i(\mathbf{W}^i)^{\top} + O_P(\tau_n^{-1}) \\
&= n^{-1}\sum_{i=1}^{n}\left(K^2\right)_{\mathbf{h}}(\mathbf{x},\mathbf{X}^i) I(\mathbf{Z}^i = \mathbf{z})\mathbf{I}_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{z}) \otimes \mathbf{W}^i(\mathbf{W}^i)^{\top} + O_P(\tau_n^{-1} + h_{\max}^2 + \lambda_{\max}) \\
&= \boldsymbol{\Sigma}_{\infty} + o_P(1)
\end{aligned} \tag{A.20}$$

Note that the last approximation in (A.20) can be obtained in our conditions of αmixing coefficients by applying arguments in the proof of Theorem 2.1. Fix a subsequence $(n') \subset \mathbb{N}$.

By (A.19) and (A.20), we have a further subseqnece $(n'') \subset (n')$ such that

$$\forall \varepsilon > 0, \ \sum_{i=1}^{n} E^* \left[\left\|\boldsymbol{\chi}^{i,*}\right\|^2 I\left(\left\|\boldsymbol{\chi}^{i,*}\right\| > \varepsilon\right)\right] \to 0 \text{ a.s. along } (n'')$$

$$\mathrm{Var}^* \left((nh_1 \times \cdots \times h_d)^{1/2} \hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) \right) \to \boldsymbol{\Sigma}_\infty \text{ a.s. along } (n'')$$

By Lindeberg central limit theorem, we get for all $\mathbf{w} \in \mathbb{R}^{d+1}$

$$P^* \left((nh_1 \times \cdots \times h_d)^{1/2} \hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) \le \mathbf{w} \right) \to P\left(N(\mathbf{0}, \boldsymbol{\Sigma}_\infty) \le \mathbf{w}\right) \text{ a.s. along } (n'') \tag{A.21}$$

where P^* denotes the conditional probability given $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$, $1 \le i \le n$. Since the distribution function of $N(\mathbf{0}, \boldsymbol{\Sigma}_\infty)$ is continuous, by Polya's theorem, (A.21) implies

$$d_{\mathrm{K}} \left(\mathcal{L}^* \left((nh_1 \times \cdots \times h_d)^{1/2} \hat{\mathbf{F}}^*(\mathbf{x}, \mathbf{z}) \right), N(\mathbf{0}, \boldsymbol{\Sigma}_\infty) \right) \to 0 \text{ a.s. along } (n'')$$

Therefore, we get (A.18) and Theorem 3.1.

Now, we prove Theorem 3.2. As in the case of Theorem 3.1, it is sufficient to prove (A.18). By Polya's theorem, Lindeberg central limit theorem for martingale difference arrays(Gaenssler et al. (1978)) and Wald's device, it is enough to show that

$$
{}^{\forall}\varepsilon > 0, \ \sum_{i=1}^{n} E^{*}\left[\left\|\boldsymbol{\chi}^{i,*}\right\|_2^2 I\left(\left\|\boldsymbol{\chi}^{i,*}\right\|_2 > \varepsilon\right) | \mathcal{G}_{n,i-1}^{*}\right] = o_P(1), \tag{A.22}
$$

$$
\sum_{i=1}^{n} E^{*}[\boldsymbol{\chi}^{i,*}(\boldsymbol{\chi}^{i,*})^{\top} | \mathcal{G}_{n,i-1}^{*}] = \boldsymbol{\Sigma}_{\infty} + o_P(1) \tag{A.23}
$$

where $\mathcal{G}_{n,i}^{*}$ is the σ-field generated by $\{Y^{1,*}, \ldots, Y^{i,*}\}$. In a similar way to the proof of (A.19), we can prove (A.22).

To show (A.23), we claim that

$$
n^{-1} \sum_{i=1}^{n} (K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i) \left[w_{\mathrm{d}}^{i}(\mathbf{z})^2 - w_{\mathrm{d}}^{i,*}(\mathbf{z})^2\right] = o_P(1) \tag{A.24}
$$

We first prove the case of $d_l = 1$, i.e. $\mathbf{Z}_2^i = Y^{i-1}$. Without loss of generality, assume $Y^i \in \{1, 2, \ldots, d_Y\}$. Define $p_j^i = P\left(Y^i = j | \mathbf{X}^i, \ldots, \mathbf{X}^1, \mathbf{Z}_1^i, \ldots, \mathbf{Z}_1^1\right)$. Note that by (C1*) and (C2*) we have $\left(p_1^i, \ldots, p_{d_Y}^i\right)^{\top} = \widetilde{\mathbf{M}}^i \cdot \left(p_1^{i-1}, \ldots, p_{d_Y}^{i-1}\right)^{\top}$ where $\widetilde{\mathbf{M}}^i$ is the $d_Y \times d_Y$-dimensional matrix whose (j,k) entry is $f(j, \boldsymbol{\theta}(\mathbf{X}^i, k))$. Since $\sum_{j=1}^{d_Y} p_j^i = 1$, we have for $i \geq 2$

$$
\mathbf{p}^i = \mathbf{M}^i \mathbf{p}^{i-1} + \mathbf{r}^i \tag{A.25}
$$

where $\mathbf{p}^i = \left(p_1^i, \ldots, p_{d_Y - 1}^i\right)^{\top}$, $\mathbf{r}^i = [f(1, \boldsymbol{\theta}(\mathbf{X}^i, d_Y)), \ldots, f(d_Y - 1, \boldsymbol{\theta}(\mathbf{X}^i, d_Y))]^{\top} \in \mathbb{R}^{d_Y - 1}$ and $\mathbf{M}^i$ is the $(d_Y - 1) \times (d_Y - 1)$-dimensional matrix whose (j,k) entry is $f(j, \boldsymbol{\theta}(\mathbf{X}^i, k)) - f(j, \boldsymbol{\theta}(\mathbf{X}^i, d_Y))$. Analogously, define $p_j^{i,*} = P^{*}\left(Y^{i,*} = j\right)$ and $\mathbf{p}^{i,*} = (p_1^{i,*}, \ldots, p_{d_Y-1}^{i,*})^{\top}$. $\widehat{\mathbf{M}}^i$ and $\hat{\mathbf{r}}^i$ are defined by replacing $\boldsymbol{\theta}$ in the expression of $\mathbf{M}^i$ and $\mathbf{r}^i$ by $\hat{\boldsymbol{\theta}}$, respectively. We also have for $2 \leq i \leq n$

$$
\mathbf{p}^{i,*} = \widehat{\mathbf{M}}^i \mathbf{p}^{i-1,*} + \hat{\mathbf{r}}^i. \tag{A.26}
$$

Combining (A.25) and (A.26), we obtain for $2 \leq i \leq n$

$$\boldsymbol{\Delta}^i = \mathbf{M}^i \boldsymbol{\Delta}^{i-1} + \boldsymbol{\varepsilon}^i \tag{A.27}$$

where $\boldsymbol{\Delta}^i = \mathbf{p}^i - \mathbf{p}^{i,*}$ and $\boldsymbol{\varepsilon}^i = \left[\mathbf{M}^i - \widehat{\mathbf{M}^i}\right] \mathbf{p}^{i-1,*} + \mathbf{r}^i - \hat{\mathbf{r}}^i$.

Let $\mathbf{M}^{[j:k]} = \mathbf{M}^k \times \cdots \times \mathbf{M}^j$ if $j \leq k$, and identity matrix otherwise. From (A.27), we get for $2 \leq i \leq n$

$$\boldsymbol{\Delta}^i = \mathbf{M}^{[2:i]} \boldsymbol{\Delta}^1 + \sum_{j=2}^{i} \mathbf{M}^{[(j+1):i]} \boldsymbol{\varepsilon}^j. \tag{A.28}$$

Let $\|\mathbf{A}\|_{\max} = \max_{j,k} |a_{jk}|$ for a matrix $\mathbf{A} = (a_{jk})$ and $C_2 = (d_Y - 1)(1 - d_Y C_1)$. From (C3*), $0 \leq C_2 < 1$ and we can show that

$$\left\| \prod_{j=1}^{k} \mathbf{B}_j \right\|_{\max} \leq C_2^k \text{ for } k \geq 1, \ \mathbf{B}_j \in \left\{ \mathbf{M}^i : i \geq 2 \right\} \tag{A.29}$$

by induction. (A.28) and (A.29) yield that for $2 \leq i \leq n$

$$\left\| \boldsymbol{\Delta}^i \right\|_2 = O_P \left(C_2^{i-1} + \tau_n^{-1} \sum_{j=2}^{i} C_2^{i-j} \right) = O_P \left(C_2^{i-1} + \tau_n^{-1} \frac{1 - C_2^{i-1}}{1 - C_2} \right) \tag{A.30}$$

Note that we also have used $\|\boldsymbol{\varepsilon}^i\|_2 = O_P(\tau_n^{-1})$ (from Theorem 2.1) to derive (A.30). Since $\max_{1 \leq i \leq n} |w_{\mathrm{d}}^i(\mathbf{z})| \leq 1$, $\max_{1 \leq i \leq n} |w_{\mathrm{d}}^{i,*}(\mathbf{z})| \leq 1$, the variance part of the left hand side of (A.24) is negligible. Applying (A.30), we have (A.24) from

$$
\begin{aligned}
&\left| n^{-1} \sum_{i=1}^{n} E\left[(K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i) \left\{ w_{\mathrm{d}}^{i}(\mathbf{z})^2 - w_{\mathrm{d}}^{i,*}(\mathbf{z})^2 \right\} \right] \right| \\
&= \left| n^{-1} \sum_{i=1}^{n} E\left[(K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i) \left\{ I(\mathbf{Z}^i = \mathbf{z}) - I(\mathbf{Z}^{i,*} = \mathbf{z}) \right\} \right] \right| + O(\lambda_{\max}^2) \\
&\leq n^{-1} \sum_{i=2}^{n} E\left[(K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i) \left\| \boldsymbol{\Delta}^i \right\|_2 \right] + O(\lambda_{\max}^2) \\
&= O\left(n^{-1} + \tau_n^{-1} + \lambda_{\max}^2 \right)
\end{aligned}
$$

Next, we prove the case of $d_l > 2$. For notational simplicity, we show only the case Y^i is binary, $\mathbf{Z}^i = (Y^{i-1}, Y^{i-2})$ $(d_l = 2)$.
Define $p_{jk}^i = P\left(Y^i = j, Y^{i-1} = k | \mathbf{X}^i, \ldots, \mathbf{X}^1, \mathbf{Z}_1^i, \ldots, \mathbf{Z}_1^1\right)$. By (C1*) and (C2*), we have

$$
\begin{aligned}
&\left(p_{11}^i, p_{10}^i, p_{01}^i, p_{00}^i\right)^\top \\
&= \begin{pmatrix}
f(1, \boldsymbol{\theta}(\mathbf{X}^i, 1, 1)) & f(1, \boldsymbol{\theta}(\mathbf{X}^i, 1, 0)) & 0 & 0 \\
0 & 0 & f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 1)) & f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0)) \\
f(0, \boldsymbol{\theta}(\mathbf{X}^i, 1, 1)) & f(0, \boldsymbol{\theta}(\mathbf{X}^i, 1, 0)) & 0 & 0 \\
0 & 0 & f(0, \boldsymbol{\theta}(\mathbf{X}^i, 0, 1)) & f(0, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0))
\end{pmatrix}
\begin{pmatrix} p_{11}^{i-1} \\ p_{10}^{i-1} \\ p_{01}^{i-1} \\ p_{00}^{i-1} \end{pmatrix}
\end{aligned}
$$

Since $p_{11}^i + p_{10}^i + p_{01}^i + p_{00}^i = 1$, we have $\mathbf{p}^i = \mathbf{M}^i \mathbf{p}^{i-1} + \mathbf{e}^i$, $i \geq 2$ where $\mathbf{p}^i = (p_{11}^i, p_{10}^i, p_{01}^i)^\top$ and in this case

$$
\mathbf{M}^i = \begin{pmatrix}
f(1, \boldsymbol{\theta}(\mathbf{X}^i, 1, 1)) & f(1, \boldsymbol{\theta}(\mathbf{X}^i, 1, 0)) & 0 \\
-f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0)) & -f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0)) & f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 1)) - f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0)) \\
f(0, \boldsymbol{\theta}(\mathbf{X}^i, 1, 1)) & f(0, \boldsymbol{\theta}(\mathbf{X}^i, 1, 0)) & 0
\end{pmatrix}
$$

and $\mathbf{e}^i = (0, f(1, \boldsymbol{\theta}(\mathbf{X}^i, 0, 0)), 0)^\top$.

Similarly, define $p_{jk}^{i,*} = P^*(Y^{i,*} = j, Y^{i-1,*} = k)$ and $\mathbf{p}^{i,*} = (p_{11}^{i,*}, p_{10}^{i,*}, p_{01}^{i,*})^\top$. $\widehat{\mathbf{M}}^i$ and $\hat{\mathbf{e}}^i$ are defined by replacing $\boldsymbol{\theta}$ in the expression of $\mathbf{M}^i$ and $\mathbf{e}^i$ by $\hat{\boldsymbol{\theta}}$, respectively. We also have $\mathbf{p}^{i,*} = \widehat{\mathbf{M}}^i \mathbf{p}^{i-1,*} + \hat{\mathbf{e}}^i$, $2 \le i \le n$. Then, we can get (A.24) in a same way to the case of $d_l = 1$.

In a similar way to the proof of (A.20) with (A.24), we obtain (A.23) by

$$
\begin{aligned}
&\sum_{i=1}^{n} E^*[\boldsymbol{\chi}^{i,*}(\boldsymbol{\chi}^{i,*})^\top | \mathcal{G}_{n,i-1}] \\
&= n^{-1} \sum_{i=1}^{n} (K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i)(w_{\mathrm{d}}^{i,*})^2 \\
&\quad \times E^* \left[\dot{\boldsymbol{\ell}} \left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i,*}), Y^{i,*} \right) \dot{\boldsymbol{\ell}} \left(\hat{\boldsymbol{\theta}}(\mathbf{X}^i, \mathbf{Z}^{i,*}), Y^{i,*} \right)^\top \middle| \mathcal{G}_{n,i-1} \right] \otimes \mathbf{W}^i (\mathbf{W}^i)^\top \\
&= n^{-1} \sum_{i=1}^{n} (K^2)_{\mathbf{h}}(\mathbf{x}, \mathbf{X}^i)(w_{\mathrm{d}}^{i})^2 \\
&\quad \times E \left[\dot{\boldsymbol{\ell}} \left(\boldsymbol{\theta}(\mathbf{X}^i, \mathbf{Z}^{i}), Y^{i} \right) \dot{\boldsymbol{\ell}} \left(\boldsymbol{\theta}(\mathbf{X}^i, \mathbf{Z}^{i}), Y^{i} \right)^\top \middle| \mathbf{X}^i, \mathbf{Z}^i \right] \otimes \mathbf{W}^i (\mathbf{W}^i)^\top + o_P(1) \\
&= \boldsymbol{\Sigma}_\infty + o_P(1)
\end{aligned}
$$

From (A.22) and (A.23), we get (A.18) and this completes the proof of Theorem 3.2.

References

Aerts, M. and Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association*, 92(440):1536–1545.

Bhattacharya, R. and Rao, R. (1986). *Normal Approximation and Asymptotic Expansions*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *The Michigan Mathematical Journal*, 30(1):69–81.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179.

Claeskens, G. and Keilegom, I. v. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics*, 31(6):1852–1884.

Deimling, K. (1985). *Nonlinear Functional Analysis*. Springer-Verlag Berlin Heidelberg.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141–150.

Franke, J., Kreiss, J.-P., and Mammen, E. (2002a). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*, 8(1):1–37.

Franke, J., Kreiss, J.-P., Mammen, E., and Neumann, M. H. (2002b). Properties of the nonparametric autoregressive bootstrap. *Journal of Time Series Analysis*, 23(5):555–585.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9(6):1218–1228.

Gaenssler, P., Strobel, J., and Stute, W. (1978). On central limit theorems for matingale triangular arrays. *Acta Mathematica Academiae Scientiarum Hungariae*, 31:205–216.

Gonçalves, S. and White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119(1):199–219.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and their Applications*, 20(2):231–246.

Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, 22(2):215–232.

Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics*, 20(2):695–711.

Hall, P., Heyde, C., Birnbaum, Z., and Lukacs, E. (2014). *Martingale Limit Theory and Its Application*. Communication and Behavior. Elsevier Science.

Hall, P. and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921.

Hardle, W. and Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401):102–110.

Härdle, W., Huet, S., Mammen, E., and Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Econometric Theory*, 20(2):265–300.

Härdle, W., Mammen, E., and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, 93(444):1461–1474.

Kauppi, H. and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Liu, R. Y. and Singh, K. (1992). *Exploring the Limits of Bootstrap*, chapter Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. Wiley Series in Probability and Statistics. Wiley.

Masry, E. (1996). Multivariate local polynomial regression for time series:uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24(2):165–179.

McMurry, T. L. and Politis, D. N. (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statistics & Probability Letters*, 78(15):2463 – 2469.

Neumann, M. H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *The Annals of Statistics*, 23(6):1937–1959.

Neumann, M. H. (1997). Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure. *Statistics*, 29(1):1–36.

Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics*, 26(4):1570–1613.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9(4):307–333.

Paparoditis, E. and Politis, D. N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Annals of the Institute of Statistical Mathematics*, 52(1):139–159.

Park, B., Simar, L., and Zelenyuk, V. (2015). Categorical data in local maximum likelihood: theory and applications to productivity analysis. *Journal of Productivity Analysis*, 43(2):199–214.

Park, B. U., Simar, L., and Zelenyuk, V. (2017). Nonpa metric estimation of dynamic discrete choice models for time series data. *Computational Statistics and Data Analysis*, 108:97–120.

Park, B. U., Simar, L., and Zelenyuk, V. (2019). Forecasting of recessions via dynamic probit for time series: replication and extension of kauppi and saikkonen (2008). *Empirical Economics*.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99 – 130.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(4):797–811.

국문초록

본 연구에서는 비모수 동적 이산 선택 모형에서 붓스트랩에 대해 다루었다. 해당 모형은 반응 변수가 이산형이며 공변량으로 이산형과 연속형을 모두 포함할 수 있는 비모수 모형이다. 뿐만 아니라 공변량으로 반응 변수의 이전 시점의 값도 포함할 수 있는 시계열 모형이다. 우선 기존의 국소 가능도 추정 방법을 확장하여 비모수 동적 이산 선택 모형에서 미지의 함수를 추정하기 위한 방법론을 제안하고 해당 추정량의 여러가지 성질들을 밝힌다. 이를 바탕으로 두가지 붓스트랩 방법을 제안한다. 두가지 붓스트랩 방법의 가장 큰 차이점은 현재 시점의 반응 변수를 재생성 할 때 이전 시점에 재생성된 반응 변수 값의 활용 여부이다. 또한 재생성한 표본으로부터 일단계 붓스트랩 추정량을 정의한다. 우리는 이 두가지 붓스트랩 방법이 모두 일치성을 가짐을 증명한다. 모의 실험에서는 붓스트랩을 이용한 신뢰 구간의 성능을 살펴본다. 또한 동적 이진 프로빗 모형을 활용해 미국의 경기 침체 자료를 분석하여 회귀함수의 점별 신뢰구간을 얻는다.

주요어 : 이산 반응 모형, 비모수 모형, 붓스트랩, 시계열 자료

학　번 : 2012-20235